



Cite this: *J. Anal. At. Spectrom.*, 2022, **37**, 2672

Can deep learning assist automatic identification of layered pigments from XRF data?†

Bingjie Jenny Xu,^a Yunan Wu,^a Pengxiao Hao,^{ab} Marc Vermeulen,^{ac} Alicia McGeachy,^{ad} Kate Smith,^e Katherine Eremin,^e Georgina Rayner,^{de} Giovanni Verri,^f Florian Willomitzer,^{ag} Matthias Alfeld,^h Jack Tumblin,^a Aggelos Katsaggelos^a and Marc Walton^{id, *ai}

X-ray fluorescence spectroscopy (XRF) plays an important role for elemental analysis in a wide range of scientific fields, especially in cultural heritage. XRF imaging, which uses a raster scan to acquire spectra pixel-wise across artworks, provides the opportunity for spatial analysis of pigment distributions based on their elemental composition. However, conventional XRF-based pigment identification relies on time-consuming elemental mapping facilitated by the interpretation of measured spectra by experts. To reduce the reliance on manual work, recent studies have applied machine learning techniques to cluster similar XRF spectra in data analysis and to identify the most likely pigments. Nevertheless, it is still challenging to implement automatic pigment identification strategies to directly tackle the complex structure of real paintings, e.g. pigment mixtures and layered pigments. In addition, pigment identification based on XRF on a pixel-by-pixel basis remains an obstacle due to the high noise level. Therefore, we developed a deep-learning based pigment identification framework to fully automate the process. In particular, this method offers high sensitivity to the underlying pigments and to the pigments present in low concentrations, therefore enabling robust mapping of pigments based on single-pixel XRF spectra. As case studies, we applied our framework to lab-prepared mock-up paintings and two 19th-century paintings: Paul Gauguin's *Poèmes Barbares* (1896) that contains layered pigments with an underlying painting, and Paul Cezanne's *The Bathers* (1899–1904). The pigment identification results demonstrated that our model achieved comparable results to the analysis by elemental mapping, suggesting the generalizability and stability of our model.

Received 17th July 2022
 Accepted 25th October 2022

DOI: 10.1039/d2ja00246a
rsc.li/jaas

1 Introduction

X-ray fluorescence spectroscopy (XRF) is a well-established workhorse technique for elemental analysis in a wide range of scientific fields,¹ such as geochemistry,^{2–4} forensic science^{5,6} and archaeology.⁷ Few areas of research benefit from its use as much as the investigation of cultural heritage that often necessitates *in situ* investigations that take place under ambient conditions

as is offered by open-architecture and hand-held versions of the XRF instrument. To fully characterize the heterogeneous nature and complex history of many artworks, XRF often requires a high number of measurements that are best implemented in the form of raster-based imaging. To form an XRF image, the instrument moves across the surface of an object, such as a painting, while collecting spectra point-by-point that are spatially redressed to their 2-D locations.⁸ XRF imaging allows for spatial analysis of pigment distributions based on their elemental composition.

Conventional XRF-based pigment identification uses spectrum evaluation methods⁹ to generate elemental maps, which existing XRF analysis software, for example, PyMCA,¹⁰ can fully support. However, identifying the pigments that cause these elemental maps and their spectra requires input from experts that have prior knowledge of the painting technique. Moreover, many artworks consist of varying pigment mixtures layered in complicated stratigraphies.^{11–13} To assist the manual work of evaluating spectra, machine learning techniques have recently been applied to pigment identification by clustering pigment-related spectral features.¹⁴ For example, XRFast, an open-

^aNorthwestern University, 2145 Sheridan Road, Evanston, IL, USA

^bShanghai University, 333 Nanchen Rd, Shanghai, China

^cThe National Archives, Bessant Dr, Richmond TW9 4DU, UK

^dThe Metropolitan Museum of Art, 1000 5th Ave, New York, NY, USA

^eHarvard Art Museums, Straus Center for Conservation and Technical Studies, 32 Quincy St, Cambridge, MA, USA

^fArt Institute of Chicago, 111 S Michigan Ave, Chicago, IL, USA

^gUniversity of Arizona, 1630 E University Blvd, Tucson, AZ, USA

^hDelft University of Technology, 2628 CN Delft, Netherlands

ⁱM+, 38 Museum Drive, West Kowloon Cultural District, Hong Kong, China. E-mail: marc.walton@mplus.org.hk

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d2ja00246a>

source unsupervised sparse dictionary learning algorithm developed recently by our group, finds maps of correlated elements to help in pigment identification, which is an improvement over the traditional approach that calculates image correlations.¹⁵ Today, deep learning (DL) has been widely applied to assist with XRF analysis and has the potential to perform fully automatic identification of elements and their sources. For example, Shugar *et al.* identified 48 different wood species with the XRF dataset using convolutional neural networks, reaching an accuracy of 99%.¹⁶ They found the range between 0.7–1.7 keV the most important portion of the spectra for wood classification, which covers the elements calcium, aluminum and magnesium. Moreover, Kim *et al.* applied a neural network on micro XRF data to generate mineral maps on natural rocks. They showed that DL was a good way to improve the description of mineral reactivity to rock samples of different origin, size, and thickness.¹⁷ Most recently, Jones *et al.* proposed a deep-learning-based method to directly identify pigments from XRF spectra.¹⁸ By training a convolutional neural network, they classified XRF spectra into one of the 15 pigment classes with an high accuracy, but claimed that it was still challenging to apply the model to more complex scenarios, including layered pigments and pigment mixtures. Therefore, focusing on the complicated stratigraphies of real paintings, this work builds on these previous studies by proposing a convolutional neural networks to automatically identify pigment mixtures in layered structures and to display 2D pigment maps based on the probability of their occurrence.

Here we propose an end-to-end pigment identification framework, including pigment library creation, XRF spectra simulation, mock-up preparation, a pigment identification DL model, and 2D pigment map generation. As a case study, we applied our framework to a 19th-century painting, Paul Gauguin's *Poèmes Barbares* (1896), focusing on a set of 19-century pigments previously identified in this painting.¹⁹ In addition, previous analysis revealed a hidden painting beneath the surface, which tremendously increased the difficulty of pigment interpretation. Therefore, our DL model targets pigment identification in the multilayered matrices of the painting. By training the DL model using 16 224 simulated XRF spectra of three-layered pigments, followed by finetuning of the model using 20% of the experimental XRF spectra (*i.e.*, 1320 XRF spectra) from mock-ups, the DL model demonstrated satisfying performance of pigment identification on the mock-ups as well as the painting *Poèmes Barbares*. In particular, a high sensitivity toward identifying pigments present in low concentrations is shown.

To further demonstrate the applicability of this approach, we applied the finetuned model to Paul Cezanne's *The Bathers* (1899–1904). This is a single-layered painting created from a comparable time period to *Poèmes Barbares*, composed of similar but fewer pigments. The DL model achieved high probabilities in identifying the pigments in *The Bathers*, suggesting our model's generalizability and stability.

In all, our framework provides an automatic and quick pigment identification strategy based on non-invasive XRF imaging, in particular targeting the paintings' complex layered

structure to the XRF response. The trained model does not require expertise or extensive familiarity working with XRF and pigments, but directly answers where the pigment might exist. Although the type of pigments are limited to those trained in the current work, our framework shows great potential for extension to other types of pigments and paintings, as well as XRF-based identification problems in the fields beyond cultural heritage.

2 Experimental method

XRF data from gathered from three sources to build, train, and test our framework: from existing paintings, from oil-paint 'mock-ups' of crossed paint stripes of different pigment mixtures, and simulated spectra of multilayered pigments.

2.1 XRF datasets

Paul Gauguin's *Poèmes Barbares* (1896) was selected to test our pigment identification approach. This painting is representative of the challenging pigment identification tasks present in many 19th-century paintings, which often involve a plethora of pigments made available after the industrial revolution. In addition, the types of pigments and the structure of the paint layers of *Poèmes Barbares* were studied previously by combining XRF, reflectance imaging spectroscopy and cross-section analysis,¹⁹ therefore providing us with reliable ground-truth measurements of this historical work.

These previous studies show that *Poèmes Barbares* consists of a visible (top) painting and a hidden (bottom) painting, each constructed with multiple paint layers and various mixtures of pigments.¹⁹ To accommodate this complex layer structure for the DL approach, we simplified it into a three-layer structure: one top pigment layer, one bottom pigment layer, and one ground layer. Referring to the chemical analysis of multiple cross-sections of the painting, we chose 11 distinct pigments of interest (Table 1) and calcium carbonate as the ground layer.

Next, we made our own set of layered oil paint mock-ups to better measure how the layers affect the pigments' XRF spectra. For these layered paintings (Fig. 1), we chose a range of the pigment fractions, binder ratios, and layer thicknesses as related to *Poèmes Barbares*. We then created sets of 3-layer mock-ups consisting of crossed strips of paint with these values and prepared them for XRF measurement, generating 6605 XRF spectra including 64 pigment layer structures. Unfortunately, the experimental data caused the overfitting of the deep learning model due to the lack of variation in the pigment layer structure. To tackle this problem, we further generated a simulation dataset for these same pigments and thicknesses to train and validate the deep learning model.

2.1.1 Experimental dataset. The experimental dataset contained XRF spectra of mock-ups with known pigment layer structure and was used for training and testing the deep learning model. We prepared the three-layered mock-ups with various combinations of pigments: six pigment mixtures as the bottom paint layer (mainly varying in the pigments' mass fractions, providing 16 bottom layers in total) and four mixtures as

Table 1 Pigment library for selected pigments

Index	Pigment ^a	Chemical formula
1	Calcium carbonate	CaCO_3
2	Chrome oxide green (CrG)	Cr_2O_3
3	Chrome yellow (CrY)	PbCrO_4
4	Cobalt blue (CB)	$\text{CoO}\cdot\text{Al}_2\text{O}_3$
5	Emerald green (EG) ^b	$\text{Cu}(\text{CH}_3\text{COO})_2\cdot 3\text{Cu}(\text{AsO}_2)_2$
6	Iron oxide (IO)	Fe_2O_3
7	Lead white (LW)	$2\text{PbCO}_3\cdot\text{Pb}(\text{OH})_2$
8	Prussian blue (PB)	$\text{Fe}[\text{Fe}_2(\text{CN})_6]_3$
9	Red lead (RL)	Pb_3O_4
10	Carmine (CM) ^c	SnO_2
11	Vermilion (VM)	HgS
12	Zinc white (ZW)	ZnO

^a Calcium carbonate, chrome oxide green, cobalt blue, verdigris, iron oxide red (120 M), lead white, Prussian blue, and vermillion were purchased from Kremer Pigmente (New York, NY). Chrome yellow and red lead were purchased from Rublev Colours (Willits, CA). Zinc oxide was obtained from Gamblin Artists Colors (Portland, OR). Sodium arsenite ($\geq 90\%$) and tin oxide (99.99%) were obtained from Sigma Aldrich. ^b Due to the current unavailability of commercial emerald green pigment, we mixed the copper carbonate pigment verdigris (VG, $\text{CuCH}_3\text{COO}_2\cdot 2\text{Cu}(\text{OH})_2$) and sodium arsenite (SA, NaAsO_2) to approximate the XRF signal of emerald green, in which the Cu-As mass ratio was set accordingly. ^c Since the chemical analysis suggested that tin oxide (SnO_2) was the support of carmine, we solely used the SnO_2 powder in preparing the mock-ups to represent carmine in the XRF dataset.

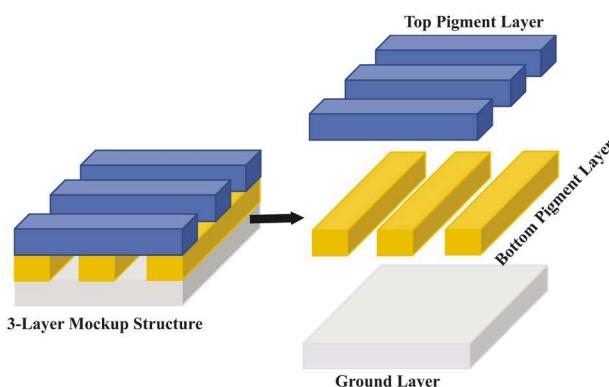


Fig. 1 The structure of mock-up samples. The first and second layers both contained multiple strips of pigment layers, each strip with its pigment combination selected from Table 1; the third layer was a ground layer consisting of calcium carbonate.

the top paint layer, making up 64 different layer structures (Table 2). Fig. 1 illustrated the structure of our mock-up samples: the first and second layer both contained multiple strips of pigment layers, where the combination of pigments in each strip was selected from Table 2; the third layer was a fixed ground layer of calcium carbonate.

A tape casting coater (model MSK-AFA-HC100, MTI Corporation (Richmond, CA)) was used to deposit the paint layer sequentially with adjustable thicknesses. A new layer was painted after the previous layer completely dried. In all mock-ups, we applied a calcium carbonate ground layer of 150–200 μm thickness to match that of Poèmes Barbares. As for the

Table 2 Summary of pigment mixtures in the mock-ups

Pigment mixture	Compound ^a	Mass fraction (%)
Top 1	LW, CM, PB, CB	55, 10, 25, 10
Top 2	VM, CM, CB	30, 30, 40
Top 3	VM, CrY, IO, VG, SA	39, 6, 39, 6, 10
Top 4	VM	100
Bottom 1A	ZW, LW, RL, VM	10, 10, 10, 70
Bottom 1B	ZW, LW, RL, VM	10, 20, 50, 20
Bottom 1C	ZW, LW, RL, VM	10, 50, 20, 20
Bottom 2A	ZW, LW, VM	10, 85, 5
Bottom 2B	ZW, LW, VM	10.7, 88.7, 0.6
Bottom 2C	ZW, LW, VM	10, 50, 40
Bottom 3A	ZW, LW, CB, CrG	10, 15, 15, 60
Bottom 3B	ZW, LW, CB, CrG	10, 15, 60, 15
Bottom 3C	ZW, LW, CB, CrG	10, 60, 15, 15
Bottom 4A	ZW, LW, CrY, VG, SA	10, 45, 15, 12, 18
Bottom 4B	ZW, LW, CrY, VG, SA	10, 30, 30, 12, 18
Bottom 4C	ZW, LW, CrY, VG, SA	10, 15, 15, 24, 36
Bottom 5A	ZW, VM, CrY, VG, SA	10, 5, 5, 32, 48
Bottom 5B	ZW, VM, CrY, VG, SA	78, 6, 6, 4, 6
Bottom 5C	ZW, VM, CrY, VG, SA	10, 15, 15, 24, 36
Bottom 6	VM	100

^a The pigments and their corresponding abbreviations used in this manuscript are listed as below: CB: cobalt blue, CrG: chrome oxide green, CM: carmine, CrY: chrome yellow, EG: emerald green, IO: iron oxide, LW: lead white, PB: Prussian blue, RL: red lead, SA: sodium arsenite, VG: verdigris, VM: vermillion, ZW: zinc white.

pigment layers, the type of mixtures simulated the palette of both the bottom and top painting in Poèmes Barbares. To include the XRF effects commonly found in layered systems and pigment mixtures, such as shielding and matrix effects, we varied the pigment fractions in the bottom paint layers (e.g. Bottom 1A, Bottom 1B, Bottom 1C in Table 2), while the top paint layer differed in its layer thickness (30–200 μm).

In preparing the mock-ups, commercial pigments and the binder were hand-ground for 10 min to obtain a uniform mixture.²⁰ Since the organic binder would not significantly affect the XRF signal, we used Galkyd Lite (Gamblin Artists Colors (Portland, OR)) for its fast-drying property. A pigment-to-binder ratio (w%/w%) of 3:1 was applied to achieve the mobility required by the tape casting coater. For lead-containing paint mixtures which appeared dryer, we added drops of Gamsol odorless mineral spirits (Gamblin Artists Colors) to further dilute the mixture. All pigment mixtures were deposited on pH-neutral art boards (Crescent (Wheeling, IL)).

To collect the experimental XRF dataset, we scanned the mock-ups with the XGLab ELIO XRF imaging spectrometer system. The XRF spectra of the mock-ups were acquired at 40 kV and 40 μA . We set the acquisition time at 10.0 s per point to increase the signal-to-noise ratio required for deep learning. The raster scan was executed using a 100 \times 100 mm motorized X-Y linear stage mount (Zaber T-LSM100A) with a step size of 1 \times 1 mm.

2.1.2 Simulation dataset. To ensure a sufficient dataset size for training our DL model, we generated a simulation dataset of 16 224 XRF spectra in total. The spectra were calculated using the matrixSpectrum function in PyMca5.PyMcaGui.physics.xrf.

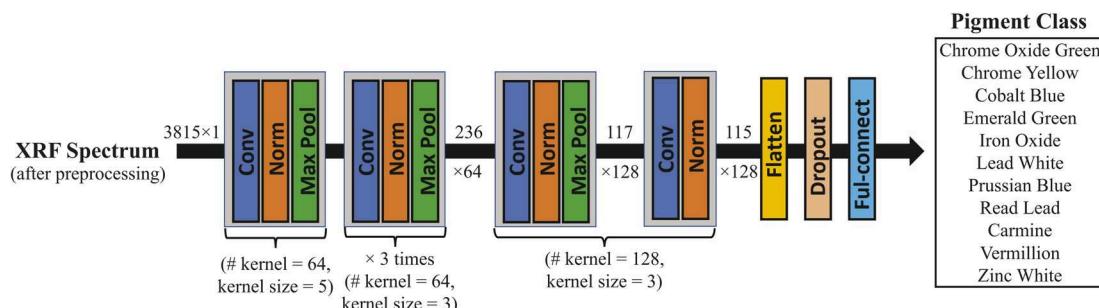


Fig. 2 The architecture of the deep learning model. It contained 1D convolutional layers (Conv), normalization layers (Norm), max pooling layers (Max Pool), a flatten layer, a dropout layer and a fully-connected layer. The input size of each XRF spectrum was 3815×1 . The output predictions were in 11 classes (one per pigment). # kernel and kernel size stood for the number and the size of the kernel at the corresponding convolutional layer, respectively.

McaAdvancedFit of the PyMCA Python package.¹⁰ These functions were used outside the PyMCA graphical user interface (GUI) in a Jupyter Notebook running Python. Basic functionality and use can be found on the NU-ACCESS Github (<https://github.com/NU-ACCESS>). Based on the fundamental parameter approach,²¹ the matrixSpectrum function simulates XRF spectra for multilayer samples. In generating the simulation dataset, we applied a three-layer structure (top pigment layer – bottom pigment layer – ground layer) similar with the mock-up paintings. Each of the top and the bottom layers consisted of a single pigment from the pigment library (Table 1), with a layer thickness of 50–200 μm (10 μm interval) and 100–150 μm (10 μm interval), respectively.

2.2 Data preprocessing

Before feeding the experimental and simulation datasets into the DL model, several preprocessing steps were necessary. First, based on preliminary ablation studies, the overlaps between the sulfur-K lines (2.31 keV) and the lead-M (2.34 keV) or mercury-M lines (2.20 keV) confused the DL model in distinguishing these elements. As a result, both the experimental and simulation

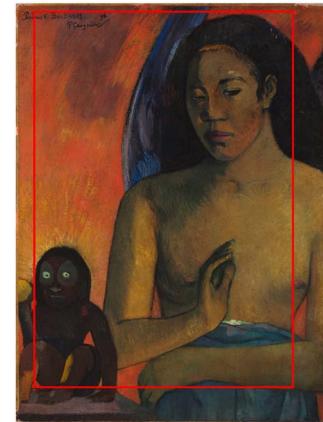


Fig. 3 "Poèmes Barbares" (1896), oil on canvas, 64.8×48.3 cm (unframed), painted by the French artist Paul Gauguin (1848–1903), Harvard Art Museums/Fogg Museum, Bequest from the Collection of Maurice Wertheim, Class of 1906. Object Number: 1951.49 © President and Fellows of Harvard College.

Table 3 The classification results for each pigment class among the models trained from three different datasets: the simulation dataset, the experimental dataset without finetuning, and the experimental dataset acquired from the mock-ups with finetuning^a

Pigment class	Simulation			Experimental (no finetune)			Experimental (finetune)		
	Accuracy	Sensitivity	F1	Accuracy	Sensitivity	F1	Accuracy	Sensitivity	F1
Cobalt blue	0.950	0.997	0.782	0.870	0.973	0.890	0.899	0.985	0.916
Emerald green	0.964	1.0	0.861	0.773	0.821	0.820	0.859	0.870	0.871
Iron oxide	0.882	0.732	0.687	0.659	0.506	0.550	0.998	0.995	0.997
Prussian blue	0.878	0.572	0.592	0.869	0.771	0.746	0.996	0.994	0.993
Carmine	1.0	1.0	1.0	0.653	0.664	0.685	0.995	0.992	0.995
Vermilion	0.987	0.997	0.955	0.918	0.994	0.952	0.947	0.978	0.970
Zinc white	0.952	1.0	0.813	0.916	0.918	0.956	0.916	0.953	0.954
Chrome yellow	0.933	0.981	0.708	0.631	0.642	0.687	0.831	0.862	0.842
Chrome oxide green	0.942	0.982	0.781	0.563	0.569	0.587	0.798	0.887	0.620
Red lead	0.871	0.653	0.626	0.508	0.458	0.556	0.868	0.687	0.634
Lead white	0.858	0.634	0.629	0.755	0.767	0.860	0.836	0.957	0.886

^a The results sum up the number of pigment class predictions of both the top- and bottom-layered pigments. The results are averaged from five-fold cross-validation.

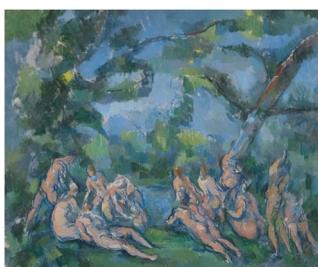


Fig. 4 The Bathers (1899–1904), oil on canvas, 51.3 × 61.7 cm, painted by the French artist Paul Cézanne (1839–1906), The Art Institute of Chicago, Amy McCormick Memorial Collection. Object Number: 1942.457 © The Art Institute of Chicago.

datasets were adonized to start at 2.80 keV to improve the performance of the DL model.

Also, the simulated XRF spectra lacked the underlying spectral background signal caused by X-ray scattering and the equipment properties. Therefore, we estimated this spectral background from the mean of the measured spectra using Statistics-sensitive Non-linear Iterative Peak-clipping (SNIP) algorithm,²² which contains a low statistics digital filter and a multipass peak clipping loop. Then we added the estimated background to the simulation dataset to better mimic their experimental counterpart, without significantly changing the key elemental peaks.

Since the XRF sensitivity to different elements varies, both the experimental and simulated spectra exhibited skewness of more than one order of magnitude, while the elemental concentrations were generally comparable. To reduce this skewness, a log–log square root transformation (eqn (1)) of the original spectrum X followed by normalization to X' was applied. This normalization step further enhanced the recognition of the elements with low peak intensities in the spectrum. X' was the final input features to the DL model.

$$X' = \log_e \left(1 + \log_e \left(1 + \sqrt{X} \right) \right), \quad (1)$$

2.3 Model architecture

The model consisted of 5 convolutional blocks, where each block was made up of a 1D convolutional layer, an activation function LeakyReLU,²³ a batch normalization layer, and a max-pooling layer, as shown in Fig. 2. According to preliminary ablation studies, the number and size of the kernels of each 1D convolutional layer were set at 64, 64, 64, 64, 128 and 5, 3, 3, 3, 3, respectively. The activation function LeakyReLU, *i.e.*, $f(x) = \max(0.01x, x)$, kept the positive part of its input while preventing “the dead ReLU issue” by using a small value when the input was negative.²³ The batch normalization layer sped up the training process by distributing the input for every layer around the same mean and standard deviation. The max-pooling layer downsampled the dimension of the input to half. Finally, the model was followed by a post-convolutional layer with 128 kernels with a size of 3, a normalization layer, a dropout layer

with a rate of 0.25, and one fully-connected layer of 11 classes, outputting the probabilities of each class (each pigment) between 0 to 1. The output layer used a sigmoid activation function $\sigma(z_j) = \frac{1}{1 + e^{-z_j}}$, where z_j was the predicted score from the model of each class. The probabilities as predicted further built the pigment maps of the paintings. Two datasets, the simulation and experimental datasets, were used to train the model and compared. The two training processes applied the same architecture as described but differed in the initial weights. The model trained with the simulation dataset used randomly initialized weights. It was then applied as a pre-trained model, its weights used as the initial value and further finetuned in training the model with the experimental dataset, a strategy known as Transfer Learning (TL).²⁴

TL is a popular technique that uses the pre-trained weights from an initial model as the starting point on another model, which reduces or eliminates the risk of overfitting and allows for better training speed and model performance.²⁵ Therefore in this paper, we pre-trained the model on the simulation dataset and then refined it using the experimental dataset. Specifically, when training with the experimental dataset, the pre-trained weights from the simulation dataset were first fixed (*i.e.*, untrainable) in all convolutional layers, whereas only the fully connected layers were finetuned with the targeted dataset. Next, all layers were trainable and were further finetuned with the experimental dataset. The model performances with and without finetuning were compared in the result section to show the effectiveness of TL.

Moreover, the loss was calculated to optimize the performance of the model by averaging the binary cross entropy of each predicted class, as defined in eqn (2).

$$\text{Loss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K [\hat{z}_j \cdot \log(z_j) + (1 - \hat{z}_j) \cdot \log(1 - z_j)], \quad (2)$$

where \hat{z} was the ground truth label, z was the score predicted from the model for each class, K was the number of the class, and N was the number of batch size.

2.4 Training strategy

The training process was completed in two steps. The model proposed in Fig. 2 was first pre-trained on the simulation dataset. A total of 16 224 simulated XRF spectra were split into a testing dataset and a training dataset with a ratio of 1 : 4. In training the model, we applied the five-fold cross-validation²⁶ by further dividing the training dataset into ‘5’ groups of equal size and iteratively selecting one group as the validation set, while the rest remaining as the training set. Therefore, with the full iteration, the model’s performance was evaluated by the testing dataset five times. All cohorts of the dataset were preprocessed following Section 2.2. The model was trained with the Adam optimizer²⁷ with an initial learning rate of 0.001. Adam was chosen for the model due to its robustness, less convergence time and fewer parameters for tuning. The batch size was 64, and on average, it took about 0.95 hours for each fold in the 5-fold cross-validation for 150 epochs with early stopping settings. Next, the experimental dataset had 6604 XRF spectra

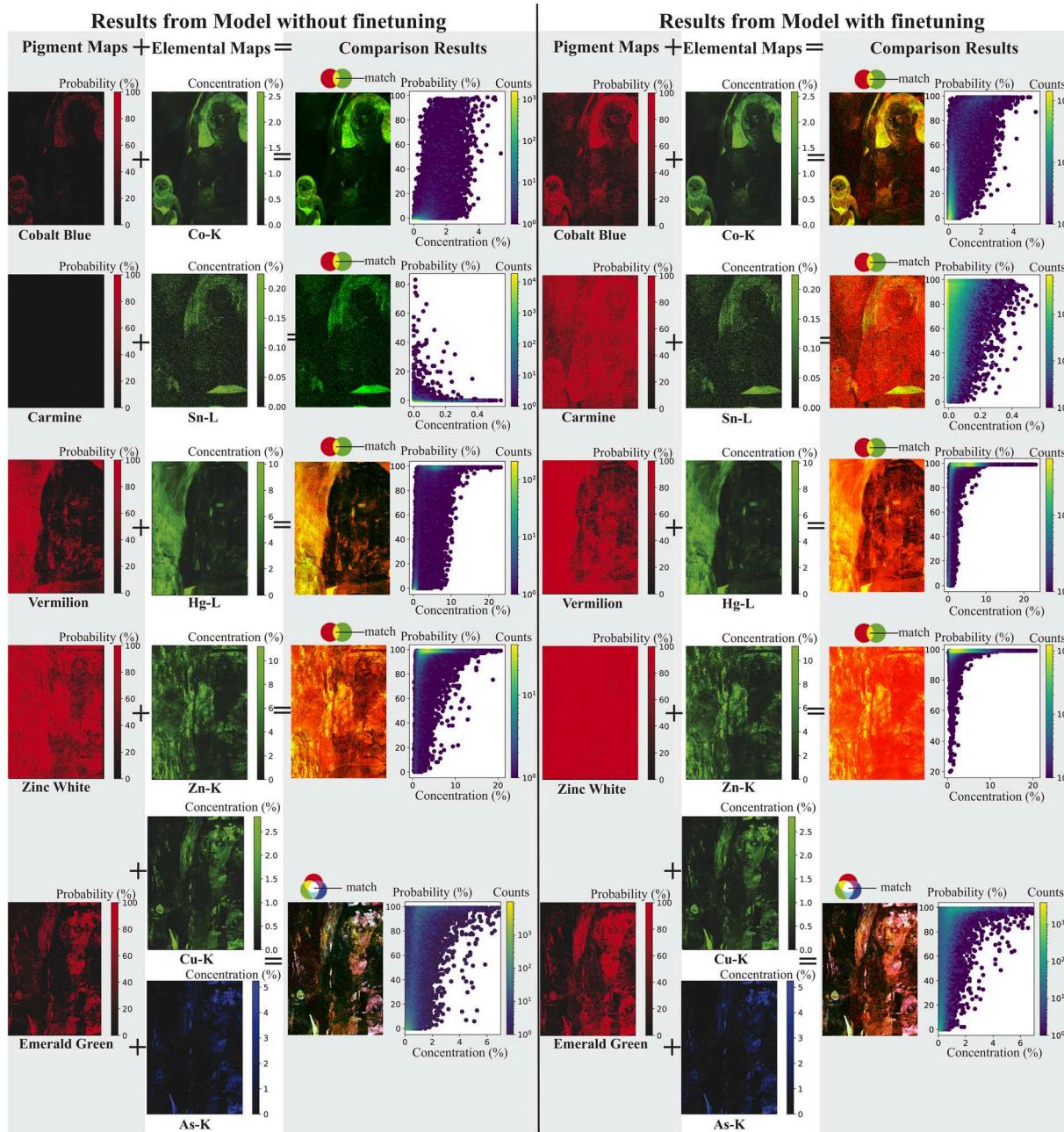


Fig. 5 Pigment identification results of Poèmes Barbares from the model without finetuning (left half), and the results from the finetuned model (right half), with results from one pigment on each row. The first column of images in each row shows the pigment map, with highest probability in red. The second column of images show the elemental map(s) for the pigment calculated by PyMCA, with highest concentrations in green or blue. The third column images overlay the first two for comparison: it combines the red pigment map with the elements maps in green and (sometimes) blue. Yellow or white areas depict strong agreement between pigment maps and their corresponding elemental maps. The fourth column scatter plot compares pigment probability and element concentration data for all image points, where element concentration sets the x-axis value (or for the bottom row, the minimum concentration of two elements), and the pigment probability sets the y-axis value. The right half of the figure shows how finetuning our DL model improves its results, and depicts images in the same arrangement used in the figure's left half.

from the two-layer pigment areas, which we manually picked from all mock-ups with known ground truth. To confirm the effect of finetuning on pigment identification, we tested the model before and after finetuning by the experimental dataset. Specifically, to test the model before finetuning, which was trained with the simulation dataset, all experimental datasets

were used as the testing dataset. The model after finetuning was initialized with the weights that performed the best among the five-fold cross-validation and was further finetuned with 20% of the randomly selected data from the experimental dataset, *i.e.* 1320 XRF spectra. The remaining 80% (5284 XRF spectra) was used as the testing dataset. Similar to the training process of the

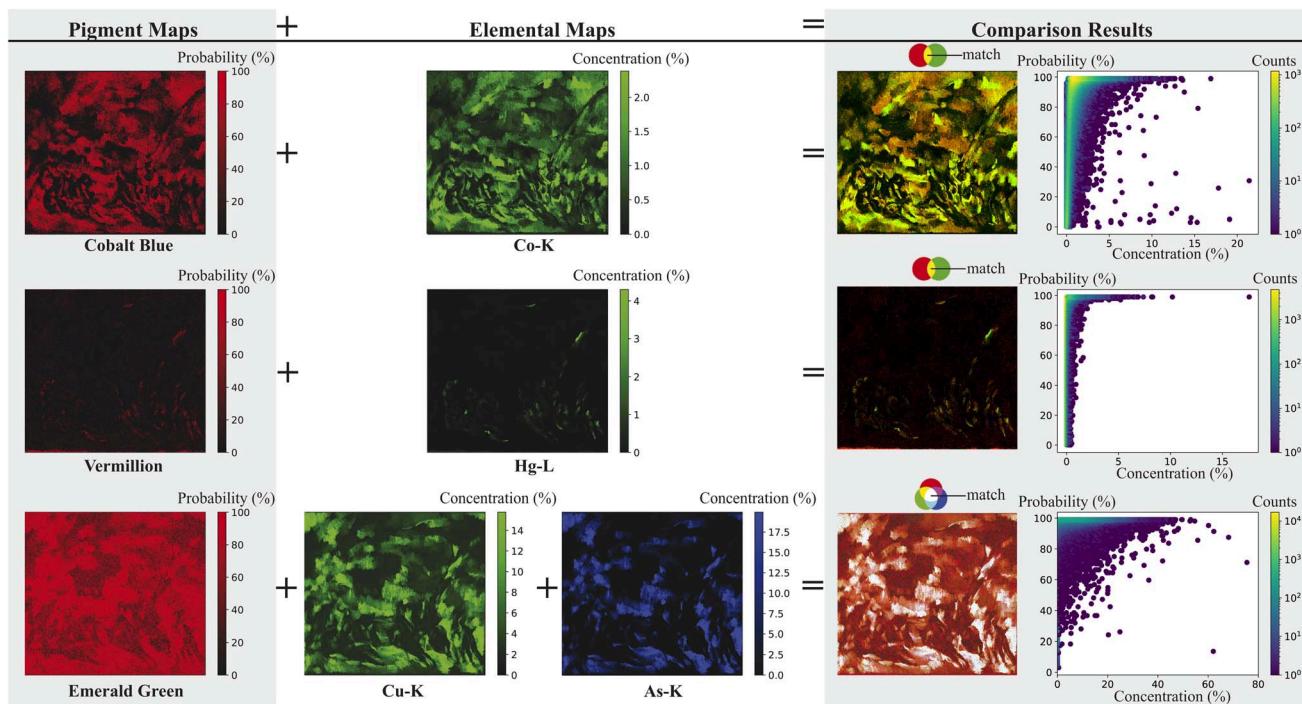


Fig. 6 Pigment identification results of The Bathers from the finetuned model, including cobalt blue, vermillion, and emerald green from top to bottom. The first, second and third columns display the pigment map, element map(s) and the comparison figure and scatter plots, respectively.

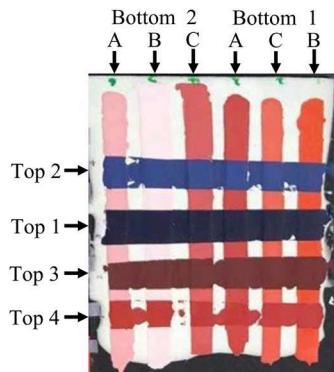


Fig. 7 Photo image of one mock-up that contains top 2, top 1, top 3, top 4 (from top to bottom) as top layers and bottom 2 with three different mass fractions (A, B and C), bottom 1 with three different mass fractions (A, C and B) (from left to right) as bottom layers.

model trained with the simulation dataset, the finetuned model was trained with the Adam optimizer with a lower learning rate of 0.0005. The batch size was 64, and it took an average of 0.2 hours for each of the 5 groups or 'folds'. All training and testing processes were performed on an NVIDIA GeForce RTX 2070 GPU using Tensorflow 2.0 in Python 3.7.

3 Results

3.1 Pigment identification model performance

Table 3 shows the effectiveness of different training approaches for pigment classification; first trained solely with simulation

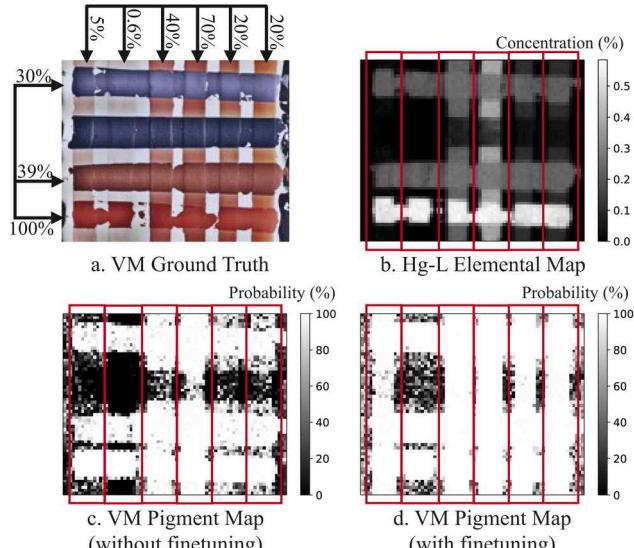


Fig. 8 Comparison between pigment maps and elemental maps for the mock-up. (a) The ground truth of the location of vermillion (VM). (b) The Hg-L element map directly generated from PyMCA. (c) The VM pigment map generated from the model without finetuning identifies Hg in the 1, 3, and 4 rows and the 3 and 4 columns. It can barely detect Hg in the 5 and 6 columns. (d) The VM pigment map generated from the finetuned model indicates VM presence in the 1, 3, and 4 rows and 1, 3, 4, 5, and 6 columns with high probabilities. The only missing column of the VM pigment map is the second column, which contains 0.6% of vermillion.

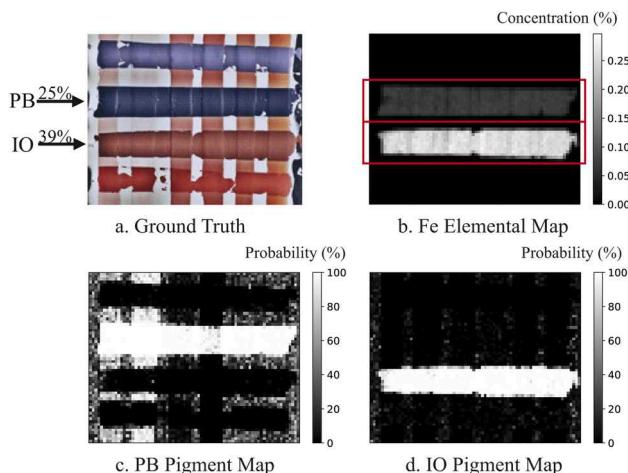


Fig. 9 Our model distinguishes Prussian blue (PB) and iron oxide (IO) in one mock-up. (a) The ground truth of PB and IO locations. (b) The Fe element map directly generated from PyMCA contains both PB and IO. (c) The PB pigment map generated from our model. (d) The IO pigment map generated from our model.

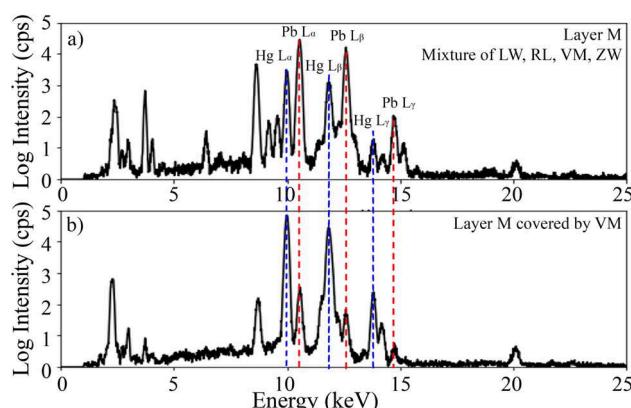


Fig. 10 XRF spectra of the mock-up painting comparing the effect of highly absorbing pigments. (a) One single layer M that contains lead white (LW), red lead (RL), vermillion (VM) and zinc white (ZW). The lead peaks marked in red have high intensities. (b) The single layer M covered by one layer of vermillion (VM). The existence of VM in the top layer significantly blocks the XRF signal of the lead-containing pigments at the bottom, challenging our model's ability to detect the hidden lead element.

data, then with experimental data without finetuning, then with the experimental data with finetuning. The overall accuracy, sensitivity, and F1 score were calculated for each class (each pigment) averaged from the five fold validation groups of experimental data on the testing datasets. Accuracy and sensitivity are defined in eqn (3) and (4).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}}, \quad (3)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (4)$$

F1 score (the harmonic mean of precision and sensitivity) evaluates the imbalanced classes, as defined in eqn (5).

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}, \quad (5)$$

where

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (6)$$

The output predictions are classified into TP, TN, FP, and FN, which are short for true positive, true negative, false positive, and false negative, respectively. True or false denotes whether the class exists or not according to the ground truth. Positive or negative suggests if the (pigment) class is predicted as existing or non-existing. All four evaluation metrics range from 0 to 1, and the closer to 1, the better the performance of the model. The analyses are performed on the same GPU using the scikit-learn package in Python 3.7.

As shown in Table 3, the model trained from the simulation dataset generally provided satisfactory accuracy, which ranges from 0.858 to 1.0. Sensitivity varies from 0.572 to 1.0, and F1 score varies from 0.592 to 1.0. According to the sensitivity and F1 score, two groups of pigments – iron-containing-only Prussian blue and iron oxide, and lead-containing-only red lead and lead white – perform worse than other pigments due to the similar elemental profiles within each group. This model can be generalized to the experimental data without finetuning but with relatively worse performance on carmine (tin-based, according the cross-section analysis), chrome oxide green, and chrome yellow. Compared with the simulation data, tin has a much lower concentration in carmine,²⁸ causing a bigger error and therefore a lower accuracy for the experimental data. In the simulation dataset, there is no pigment mixture. Chrome oxide green only contains Cr and chrome yellow contains Cr and Pb, which helps the model to separate them. But Pb also exists in other pigments in the mock-ups, which confuses the model. However, finetuning the model on the experimental data significantly improves the classification results, reaching an overall accuracy ranging from 0.798 to 0.998, sensitivity from 0.687 to 0.995, and F1 from 0.634 to 0.997. The next section will further explain our finetuning strategy.

3.2 Tests on Paul Gauguin's Poèmes Barbare

Building on the ability to identify the pigments in the mock-up samples, we applied the models to the XRF dataset obtained from Poèmes Barbare (Fig. 3), which was collected also by the XGLab ELIO XRF imaging spectrometer system.¹⁹ The red rectangle Fig. 3 marks the area of investigation.

As shown in Fig. 5, the pigment maps suggest the probability of pigments' existence as predicted by the pigment identification model, where an increased pixel brightness suggests a higher probability. The probability ranges from 0–100%, calculated directly from the XRF spectrum image. Their corresponding 2D elemental maps generated from PyMCA are also shown in Fig. 5 for comparison, where a brighter pixel suggests a higher elemental concentration. In addition, each pigment

map and its corresponding elemental map were merged to better visualize their pixel differences. Note here that as the elemental maps have a higher dynamic range than the pigment maps, we adjusted the max limit of the color bar to 99th percentile of the concentration data to balance the brightness between the two maps. To further evaluate the performance of the pigment identification model, Fig. 5 also includes a series of scatter plots, which summarize the relationship between the predicted pigment probability and its actual elemental concentration at each pixel location. With a higher accuracy in identifying the pigments, the points remain higher in the plot. If the model identifies a pigment accurately, the probability (y value) will approach to zero at zero concentration, but will rapidly increase with any non-zero elemental concentration (x value).

As shown in Fig. 5, the DL model trained and finetuned solely on the mock-ups is applicable to the paintings and simulations as well. Fig. 5 displays the results of six pigments: cobalt blue, carmine, vermillion, zinc white, and emerald green, each with a unique elemental spectral profile. By comparing the models with and without finetuning, the scatter plots suggest significant effect of finetuning on improving the sensitivity and accuracy in identifying all pigments, particularly for the situation of low elemental concentrations such as carmine. As for the remaining six pigments (chrome oxide green, chrome yellow, iron oxide, Prussian blue, red lead, lead white), the current model has not yet been able to distinguish the pigments that share the same elements (Fig. S2 and S3 in ESI†), as discussed below.

3.3 Tests on Paul Cezanne's The Bathers

To further demonstrate that our model is also applicable to other paintings, we applied our same finetuned model to the XRF dataset of Paul Cezanne's The Bathers (Fig. 4). The XRF mapping was executed with a MA-XRF system at 40 kV and 1 A, with an acquisition time of 100 ms per point and a step size of 1 mm. As shown in Fig. 6, pigment maps of cobalt blue, carmine and emerald green perfectly match with their elemental maps and achieve high probabilities with concentrations increasing. The pigment identification of this painting shows results comparable to the Gauguin painting, highlighting the generalizability and stability of our model. A complete set of the elemental maps and the pigment maps with and without finetuning are given in Fig. S4 and S6 in ESI.†

4 Discussion

4.1 Finetune the model

Finetuning is a general technique popularized in deep learning models, especially on 2D images, to take advantage of weights trained on a huge dataset for another similar but smaller dataset. This technique has shown success in many fields, such as image recognition,²⁹ medical diagnosis³⁰ and unsupervised learning.³¹ In applying this method, the model initially learned the spectral features from the large simulation dataset followed by finetuning on the limited experimental data. Table 3

suggests that finetuning, even using a small subset of the dataset, can significantly improve the performance of pigment identification. In particular for the cases of low elemental concentrations, such as carmine, the accuracy, sensitivity, and F1 score increased from 0.653 to 0.995, 0.664 to 0.992, and 0.685 to 0.995, respectively.

To better visualize the effect of finetuning, we generated 2D pigment maps for one mock-up painting (Fig. 7) as an example. As shown in Fig. 8(a), the vermillion pigment (VM) was present in three horizontal paint strips (top pigment layer) and all six vertical strips (bottom pigment layer), but at different concentrations. However, neither the Hg elemental map nor the VM pigment map detected it reliably without finetuning. While Fig. 8(b) and (c) may reveal the existence of VM at low concentrations, finetuning significantly improved the identification result of VM at low concentrations in Fig. 8(d), reaching its limit near 0.6% concentration as shown in the second vertical strip.

To our best knowledge, this work is the first to apply transfer learning to pigment classification using XRF spectra. By finetuning on only 20% randomly selected of the experimental data, we observed significant improvements in pigment identification. On one hand, it improved the model performance even with a limited training dataset. On the other hand, since mock-ups are hard to make, it releases the pressure of preparing a huge experimental dataset. However, as mentioned previously, since the pigment combination in the experimental dataset is limited, the finetuned model may overfit. Nevertheless, the finetuning technique contributes to extracting the features related to some specific pigment mixtures in the painter's palette, which helps the identification of pigments, especially in a specific painting, with similar painting styles. Therefore, this finetuning technique can be applied to many different fields using XRF spectra.

4.2 Pigments with similar elemental profiles

Three groups of pigments in our pigment library posed challenges due to similar elemental profiles: the chromium-containing group (chrome oxide green and chrome yellow), the iron-containing-only group (Prussian blue and iron oxide), and the lead-containing-only group (lead white and red lead). Although the two pigments in the chromium-containing group slightly vary in their elemental map (chrome oxide green only contains Cr and chrome yellow contains Cr and Pb), the model failed to distinguish between these two pigments. This is possibly caused by Pb, which exists almost everywhere in the painting, often mixed with other pigments (e.g. lead white).

For the other two pigment groups that share similar elemental profiles, the model before finetuning cannot distinguish them and shows low to medium probability for all areas that contain the element(s). However, the finetuned model can distinguish between those pigments when present in different pigment mixtures, but this only applies when the testing dataset contains the same pigment mixtures as the finetuning dataset. One special case in our result is that the finetune model can distinguish Prussian blue (PB) and iron oxide (IO) in our mock-ups. For example, in the mock-up in Fig. 7, PB exists in the

second row, and IO appears in the third row (Fig. 9(a)). The only element detectable by XRF in Prussian blue and iron oxide is Fe. Therefore, the Fe element map (Fig. 9(b)) shows the location of both pigments and cannot be used to separate one from the other. However, Prussian blue pigment map (Fig. 9(c)) and iron oxide pigment map (Fig. 9(e)), generated from our finetuned model, show promising results of distinguishing these two pigments. The two pigment maps both have high probabilities at the ground truth location of the pigment and fairly low probabilities at the ground truth location of the other pigment. At the same time, we notice the pigment maps incorrectly identify where the Fe element has low concentration or does not exist, especially for the Prussian blue pigment map, which may be related to the high tinting strength, and therefore low concentration where used, of Prussian blue.^{19,32,33} This situation also appears in the Gauguin and Cezanne painting results: when the pigment concentration is low, the probability might vary from 0 to 100%. We infer that this is due to the normalization process during the spectrum preprocessing. Specifically, the small peaks (low intensities) are enlarged after the normalization, enhancing both useful information and noise level. In the Fe element map (Fig. 9(b)), Fe concentration is much higher in iron oxide than in Prussian blue. Therefore, the iron oxide pigment map only has a few noisy points at the blank area (only the ground layer), while the Prussian blue pigment map shows high probabilities at the 1, 2, 5, 6 columns and the blank area.

4.3 X-ray absorbent materials

Some pigments can strongly absorb photons at energies needed for X-ray fluorescence measurements, and absorption by top layers of a painting can severely distort the spectra detected from fluorescence of elements in the bottom layers. For example, the presence of vermillion in the top layer diminishes the lead peaks of lead white and red lead in the bottom layer (Fig. 10). In particular, as the energy of the Pb-L β radiation is above the Hg-L 3 edge, it is absorbed by vermillion in the top layer, decreasing the Pb-L β to Pb-L α ratio. At the same time, the absorption gives rise to fluorescent emission from Hg-L α , increasing the Hg-L α to Hg-L β ratio. The shifts in line ratios for both elements create an extraordinary case the simulated data did not adequately describe and our model cannot distinguish from the underlying lead-containing pigments. This effect was previously observed in manual data evaluation as well.³⁴

5 Conclusions

XRF-based pigment identification problems have long required expert analysis and previous knowledge. In this paper, we pursued an automatic XRF data evaluation framework using deep learning. Our initial attempts at automatically identifying individual and overlapped pigments directly from XRF spectra show promise. While our model only tests a small number of pigments (11) and a small number of layers (2 + base) in this current stage, it automatically identified pigments in two different 19th-century paintings and in the training mock-ups

and simulations they inspired. We intend this paper to stimulate further work in deep-learning assisted XRF studies for layered-pigment identification, and prompt more discussion of their feasibility and practicality for broader uses.

We focused on a set of representative pigments identified or considered present in Paul Gauguin's *Poèmes Barbares* (1896)¹⁹ as a starting point and then created mock-ups to generate experimental datasets to capture nonlinear effects of layer structures. We added a simulation dataset to reduce the need for prohibitively tedious and difficult mock-up preparation. After data preprocessing, we trained the convolutional neural network with the simulation dataset and then finetuned it with the experimental dataset, therefore obtaining the pigment identification model.

Pigment maps are the visualization of the probability output of our model. The comparison of pigment maps and their corresponding element maps shows that our model can successfully identify pigments, especially in low concentration or in overpainted layers. However, the model still has some shortcomings: (a) the model cannot always distinguish pigments with similar elemental profiles; (b) it does not work when high absorbing pigments block the radiation emitted from the hidden layer; (c) the finetuned model needs experimental data with at least a small set of ground truth measurements to prevent limited pigment mixtures in the mock-ups from causing model overfitting and wrong predictions.

This research is still in its early stages, and there are multiple directions to extend the current work. First of all, including data from other techniques, such as spectral imaging to get molecular structures, might better differentiate pigments and compensate for shortcomings (a) and (b) which are common problems with using XRF to identify pigments. Second, we only try to identify presence or absence of pigments, and not their depth or the layers' sequence from front-to-back, an important but much more challenging problem. Third, the current pigment library is limited, and training the model with a larger range of pigments can make it more accessible to different paintings. With a sufficient pigment dataset, more advanced deep learning algorithms can be applied to further boost the pigment identification performance of the model. For example, Recurrent Neural Networks (RNNs)³⁵ typically solve problems with sequential input signals because their internal cells store the information retrieved from the previous point in time and use them to generate the next point in time. It will be interesting to try how RNNs work on XRF data with similar time series structures. In addition, some unsupervised learning methods, such as autoencoders,³⁶ have been proposed to extract latent features of signals without any ground truth labels. This is another interesting approach to try in future work that may solve the problem of limited unlabeled datasets in this research area.

Data availability

The related data and code for this paper can be found at the following Github repository: deep learning assisted XRF.

Author contributions

Conceptualization, Marc Walton, Aggelos Katsaggelos, Jack Tumblin, Florian Willomitzer, Matthias Alfeld and Bingjie (Jenny) Xu; data curation, Pengxiao Hao, Marc Vermeulen and Alicia McGeachy; formal analysis, Bingjie (Jenny) Xu, Pengxiao Hao, Yunan Wu and Marc Vermeulen; funding acquisition, Marc Walton; investigation, Bingjie (Jenny) Xu, Pengxiao Hao, Marc Vermeulen and Alicia McGeachy; methodology, Bingjie (Jenny) Xu, Yunan Wu, Pengxiao Hao, Marc Vermeulen, Matthias Alfeld, Aggelos Katsaggelos and Marc Walton; resources, Kate Smith, Katherine Eremin, Georgina Rayner and Giovanni Verri; software: Bingjie Xu, Yunan Wu, Marc Vermeulen, Alicia McGeachy and Marc Walton; writing – original draft, Bingjie (Jenny) Xu, Yunan Wu, Pengxiao Hao and Matthias Alfeld; writing – review & editing, Marc Vermeulen, Alicia McGeachy, Florian Willomitzer, Matthias Alfeld, Jack Tumblin, Aggelos Katsaggelos and Marc Walton. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The analysis of the Gauguin and Cezanne paintings is part of NU-ACCESS's broad portfolio of activities, made possible by generous support of the Andrew W. Mellon Foundation as well as supplemental support provided by the Materials Research Center, the Office of the Vice President for Research, the McCormick School of Engineering and Applied Science and the Department of Materials Science and Engineering at Northwestern University. The authors gratefully acknowledge Eme-line Pouyet and Gianluca Pastorelli (formerly NU-ACCESS) for the acquisition of the MA-XRF data on Gauguin's Poèmes Barbares.

Notes and references

- 1 C. van Hoof, J. R. Bacon, U. E. A. Fittschen and L. Vincze, *J. Anal. At. Spectrom.*, 2021, **36**, 1797–1812.
- 2 H. Rowe, N. Hughes and K. Robinson, *Chem. Geol.*, 2012, **324**, 122–131.
- 3 T. D. T. Oyedotun, *Geol. ecol. landsc.*, 2018, **2**, 148–154.
- 4 P. Sarala, *Geochem. Explor. Environ. Anal.*, 2016, **16**, 181–192.
- 5 K. Langstraat, A. Knijnenberg, G. Edelman, L. Van De Merwe, A. van Loon, J. Dik and A. van Asten, *Sci. Rep.*, 2017, **7**, 1–11.
- 6 K. Nakano, C. Nishi, K. Otsuki, Y. Nishiwaki and K. Tsuji, *Anal. Chem.*, 2011, **83**, 3477–3483.
- 7 M. S. Shackley, *X-Ray Fluorescence Spectrometry (XRF) in Geoarchaeology*, Springer, 2011, pp. 7–44.
- 8 M. Alfeld and L. de Viguerie, *Spectrochim. Acta, Part B*, 2017, **136**, 81–105.
- 9 P. J. van Espen and K. H. Janssens, *Handbook of X-Ray Spectrometry: Methods and Techniques*, Marcel Dekker, Inc., New York, NY, 1993, ch. 5, pp. 181–293.
- 10 V. Solé, E. Papillon, M. Cotte, P. Walter and J. Susini, *Spectrochim. Acta, Part B*, 2007, **62**, 63–68.
- 11 M. Alfeld and K. Janssens, *J. Anal. At. Spectrom.*, 2015, **30**, 777–789.
- 12 F. P. Romano, C. Caliri, P. Nicotra, S. Di Martino, L. Pappalardo, F. Rizzo and H. C. Santos, *J. Anal. At. Spectrom.*, 2017, **32**, 773–781.
- 13 M. Alfeld, W. De Nolf, S. Cagno, K. Appel, D. P. Siddons, A. Kuczewski, K. Janssens, J. Dik, K. Trentelman, M. Walton, *et al.*, *J. Anal. At. Spectrom.*, 2013, **28**, 40–51.
- 14 S. Kogou, L. Lee, G. Shahtahmassebi and H. Liang, *X-Ray Spectrom.*, 2021, **50**, 310–319.
- 15 M. Vermeulen, A. McGeachy, B. Xu, H. Chopp, A. Katsaggelos, R. Meyers, M. Alfeld and M. Walton, *J. Anal. At. Spectrom.*, 2022, **37**, 2130–2143.
- 16 A. N. Shugar, B. L. Drake and G. Kelley, *Sci. Rep.*, 2021, **11**, 1–10.
- 17 J. J. Kim, F. T. Ling, D. A. Plattenberger, A. F. Clarens and C. A. Peters, *Appl. Geochem.*, 2022, **136**, 105162.
- 18 C. Jones, N. S. Daly, C. Higgitt and M. R. Rodrigues, *Heritage Sci.*, 2022, **10**, 1–14.
- 19 M. Vermeulen, K. Smith, K. Eremin, G. Rayner and M. Walton, *Spectrochim. Acta, Part A*, 2021, **252**, 119547.
- 20 L. F. Sturdy, *PhD thesis*, Northwestern University, 2016.
- 21 L. De Viguerie, V. A. Sole and P. Walter, *Anal. Bioanal. Chem.*, 2009, **395**, 2015–2020.
- 22 C. Ryan, E. Clayton, W. Griffin, S. Sie and D. Cousens, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 1988, **34**, 396–402.
- 23 B. Xu, N. Wang, T. Chen and M. Li, 2015, preprint, arXiv:1505.00853, DOI: [10.48550/arXiv.1505.00853](https://doi.org/10.48550/arXiv.1505.00853).
- 24 F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong and Q. He, *Proc. IEEE*, 2020, **109**, 43–76.
- 25 M. Raghu, C. Zhang, J. Kleinberg, S. Bengio, Transfusion: Understanding transfer learning for medical imaging, *Advances in neural information processing systems 32*, 2019.
- 26 M. Stone, *J. Roy. Stat. Soc. B*, 1974, **36**, 111–133.
- 27 D. P. Kingma and J. Ba, 2014, preprint, arXiv:1412.6980, DOI: [10.48550/arXiv.1412.6980](https://doi.org/10.48550/arXiv.1412.6980).
- 28 R. Dapson, *Biotech. Histochem.*, 2007, **82**, 173–187.
- 29 H.-W. Ng, V. D. Nguyen, V. Vonikakis and S. Winkler, *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 2015, pp. 443–449.
- 30 S. Khan, N. Islam, Z. Jan, I. U. Din and J. J. C. Rodrigues, *Pattern Recogn. Lett.*, 2019, **125**, 1–6.
- 31 Y. Bengio, *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, 2012, pp. 17–36.
- 32 M. Vermeulen, A. S. O. Miranda, D. Tamburini, S. E. R. Delgado and M. Walton, *Heritage Sci.*, 2022, **10**, 1–22.
- 33 L. D. Glinsman, *et al.*, *PhD thesis*, Universiteit van Amsterdam Host, 2004.
- 34 C. Neelmeijer, I. Brissaud, T. Calligaro, G. Demortier, A. Hautojärvi, M. Mäder, L. Martinot, M. Schreiner, T. Tuurnala and G. Weber, *X-Ray Spectrom.*, 2000, **29**, 101–110.
- 35 A. Sherstinsky, *Phys. D*, 2020, **404**, 132306.
- 36 P. Baldi, *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, 2012, pp. 37–49.