PNNL-38905

A case study in contrastive learning information combination: Application to technical forensics of additive manufacturing filament source identification

November 2025

Alex Hagen
Cody Nizinski
Jessie Yaros
Charlie Hayden
Cuong Ly

U.S. DEPARTMENT of ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fox: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

PNNL-38905

A case study in contrastive learning information combination: Application to technical forensics of additive manufacturing filament source identification

November 2025

Alex Hagen
Cody Nizinski
Jessie Yaros
Charlie Hayden
Cuong Ly

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Acknowledgments

This research was supported by the Nuclear Forensics Transformational Innovation Initiative (NFTI), under the Laboratory Directed Research and Development (LDRD) Program at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830.

Acronyms and Abbreviations

NFTI	Nuclear Forensics Transformation Innovation
TF	Technical Forensics
TFA	Technical Forensics Analysis
FTIR	Fourier Transform Infrared Spectroscopy
NMR	Nuclear Magnetic Resonance
XRD	X-ray Diffraction
XRF	X-ray Fluorescence
ROC	Receiver Operating Characteristic

Contents

Acknowledgments ii
Acronyms and Abbreviations iii
Contents iv
1.0 Introduction 1
2.0 Method 2
3.0 Data 4
4.0 Results 5
5.0 References 7

Figures

Figure 1. Example reconstructions from validation sample from diKale .... **Error! Bookmark not defined.**
Figure 2. Receiver operating Characteristic (ROC) curve showing the proportion of times the same manufacturer scored above a varying threshold versus the proportion of the time a different manufacturer scored above that same threshold. Separate lines indicate different ROC curves for increasing number of modalities included. 5
Figure 3. Receiver Operating Characteristic (ROC) curve showing the proportion of times the same manufacturer scored above a varying threshold versus the proportion of the time a different manufacturer scored above that same threshold. Separate lines indicate different ROC curves for samples including the indicated modality. 6

1.0 Introduction

Combination of information from disparate data sources into a single decision is a core challenge in many fields, including the field of technical forensics. Technical forensics (TF) utilizes technical characterization of questioned samples to determine properties of that sample; these properties are then used to infer information of forensic interest, such as provenance, age, or attribution. TF is utilized in traditional forensic applications, such as the attribution of material fragments from an explosive, and in nuclear forensic applications, such as the attribution of actinides which have been interdicted out of regulatory control.

The challenge of combining information from disparate sources, described alternately by many terms including “Data Fusion” and “Data Integration”, is exacerbated in the technical forensics domain due to at least two factors: the challenge of interpreting each information source singularly, and the relatively small data set sizes available. Extensive literature exists attempting to combine technical forensics information sources, both in manual and automated processes. These attempts are often bespoke to the specific information sources (such as the bi-, tri-, or quad-isotope chart (Moody, Grant, and Hutcheon 2005)), with some emerging examples of simple early- and late- fusion (, respectively).

Simultaneous to the information combination efforts described in the previous paragraph, the field of natural language processing attempted (and largely succeeded) in combining information from multiple non-technical information sources. The ecosystem of “multi-modal” language models, which can take text and images as input, and generate text and images as output, became large and diverse by 2025 (Khan et al. 2025). In a generalized sense, many of these methods are trained by learning neural networks which can convert raw text or images into a vector of numbers describing the text or image, hereafter called “embeddings” and the neural networks performing the conversion are called “embedders”. By using a separate embedder for text and images, finding coincident text and images (such as images with their captions), and optimizing the parameters of the embedders such that the embeddings for the text and the image are similar, the field has found a bridge between text and images (Girdhar et al. 2023). It is the contention of the authors of this report that this insight is not limited to text and images but instead can be extended to any modality which can be found coincidently. The subject of the rest of this report is the application of this method to example multi-modal technical forensic data. Some details about the data used in this report are not appropriate for this report, and are included in a companion report (PNNL-38669).

2.0 Method

Prior work applying neural network to TF analysis (TFA) showed the utility of unsupervised learning for representation of actinide samples, due to the dearth of labeled data suitable for supervised learning frameworks (Girard* et al. 2021). A modified technique from (Girard* et al. 2021) was developed for the data anticipated in TFA. In TFA, several types of data are anticipated:

- Bulk Measurements: Scalar quantities, such as density, which represent the entire sample.

- Spectra Like: One dimensional quantities, which represent the excitation of the sample against an excitation variable, such as energy or wavelength. Note that these may be measured in several different "channels" (such as in Raman spectroscopy where the excitation wavelength is changed), so they are represented with two dimensions (i.e. ["channel" ×"spectra length"]).

- Image Like: Two dimensional quantities, which represent measurements of the sample at different spatial locations projected into a single plane, usually in a raster scan. Like spectra-like, these may be measured in several different "channels" (such as red, green, and blue for visible light imagery), or they are represented with three dimensions (i.e. ["channel" ×"image height" ×"image width"]).

- Volume Like: Three dimensional quantities, which represent measurements of the sample at different spatial locations, usually in a voxel scan or tomographic reconstruction. Like image-like, these may be measured in several different "channels" (such as different energies of x-ray in x-ray computed tomography), so they are represented with four dimensions (i.e. ["channel" ×"volume depth" ×"volume height" ×"volume width"]).

Convolutional neural networks were developed to handle the types of data described above. These convolutional neural networks were trained to reconstruct their input after compression through blocks of convolutional neural network blocks, and decompression through blocks of convolutional up-sampling neural network blocks. In this way, they are trained to concentrate the information in their input into a small, single dimensional representation, called the embedding[1]. The spatial average at the last layer of convolutional blocks was used as the embedding for each measurement.

With embeddings computed for each measurement from a sample, combination of information from each modality is then possible. The approached proposed in this report is to cast all measurements of a sample as a sequence, therefore each sample can be represented by a matrix of size [number of measurements × embedding dimension][2]. This approach aligns the combination of information from each modality into sequence modeling, a topic of enormous amounts of data-scientific research. Specifically, it admits the application of Transformers

[1] The reconstruction objective chosen attempts to preserve **all** information in the input, including information which may be unrelated to the sample, such as imaging settings. Reconstruction was chosen because of its completely unsupervised nature in this setting, where labeled data collection and augmentation are challenging or expensive.

[2] assuming all embedding dimensions are chosen to be equal.

(Vaswani et al. 2017) to the combination of multimodal information. The attention mechanism in Transformers allows for the modulation of information passing between all items in the input sequence, to all items in the output sequence. Further, unlike static modulation techniques like gMLP (Liu et al. 2021) or MLP-Mixer (Tolstikhin et al. 2021), attention is dynamic in its modulation. The modulation of information mixing between input items in the sequence and the output is determined by a pattern which is itself dependent on the input items. This flexible mechanism fits all the desiderata for combination of information from multiple technical modalities, including the ability to concentrate on relevant information, ignore irrelevant information, and contextualize information in one input item with information from other input items. We use the common *class token* approach, where a custom sequence item is prepended to the embedded modalities in the sequence, the sequence (with prepended item) is passed through an encoder-only Transformer[3], and the output item at the location corresponding to the prepended item is used as the overall embedding for the sample.

The final piece for multimodal information mixing is the technique for training the Transformer from the previous paragraph. Contrastive learning is utilized, specifically the normalized temperature-scaled cross-entropy loss from (Chen et al. 2020), which is defined as

$$\mathcal{L} = -\frac{1}{N} \sum_{i \in [0 \ldots N]} \log \frac{\sum_{j \in [0 \ldots N]} \mathbb{1}(i,j) \exp(\mathrm{sim}(x_i, x_i)/\tau)}{\sum_{j \in [0 \ldots N]} \left(1 - \mathbb{1}(i,j)\right) \exp\left(\mathrm{sim}\left(x_i, x_j\right)/\tau\right)},$$

where N are the number of embeddings in a batch, $\mathrm{sim}(x, y)$ is a similarity function between x and y (we use the inner product $x^T y$), $\mathbb{1}(i, j)$ is an indicator function which returns 1 when i and j are from a *positive pair* and 0 otherwise, and τ is some temperature which can be scaled to encourage sharper distributions. While much of the literature uses augmentations to create positive pairs, recent work has found that positive pairs chosen from different measurements of the same object have benefits and require no development of realistic augmentations (Johnson, McDonald, and Tasdizen 2024). The proposed positive pair ontology in this report is any set of measurements from the same sample or a replicate sample. The negative pair ontology is anything that does not fit that description: i.e. any set of measurements from a sample which is not the same and not a replicate. Note that supersets of measurements (and subsets) are allowed as positive pairs in this ontology, that is to say that a sample which was measured through Raman spectroscopy and x-ray fluorescence could be paired with a replicate sample measured with only Raman spectroscopy.

[3] An encoder only Transformer returns a sequence the same size as the input sequence and allows the attention mechanism to attend to all items in the input for all items in the output.

3.0 Data

A collaboration with a government agency was initiated for the purpose of advising the USSS forensics lab on the best way to compare multi-modality characterization in a statistically rigorous manner. While the agency's historic forensic mission is in fluid analysis, the exemplar problem chosen for the present work is that of plastics. It was hypothesized that multi-modality characterization could inform based plastic used to create the item, which could then be used to trace the plastic origin manufacturer's records. The forensic laboratory at the agency obtained and analyzed 13 plastics from 13 different manufacturers in 2 replicates. These were characterized using FTIR, NMR, Raman, XRD, and XRF modalities. Pretraining data was also obtained for the training of modality specific reconstructions. These were obtained for each different modality: for Raman spectroscopy from RamanSpy (Georgiev et al. 2024), for x-ray diffraction from opXRD (Hollarek et al. 2025), for x-ray fluorescence using several sources (Manni and Viganò 2023; Kabiri et al. 2024; Von Konrat 2022), for Nuclear magnetic resonance using NP-MRD (Wishart et al. 2022), and for fourier-transform infrared spectroscopy using FTIR-Plastics (Villegas Camacho et al. 2024).The data was split into a training and validation set. The validation set included samples from five manufacturers which were in replicates 2 or 3. The training set included all other samples.

4.0 Results

The autoencoder framework was highly successful in reconstructing both the pretraining, training, and validation data. Overall loss, at the end of 1.97×10^{4} iterations, was 3.03×10^{-4}

The overall performance of the framework was moderate. When used to embed samples from a manufacturer, the closest sample to that embedding was the same manufacturer 50% of the time. An embedding from the same manufacturer was in the top 3 closest embeddings 80% of the time. Figure 2 shows that the performance was best when all five modalities were included, and Figure 3 shows that samples which included information from Raman spectroscopy was most useful for embedding the entire sample.

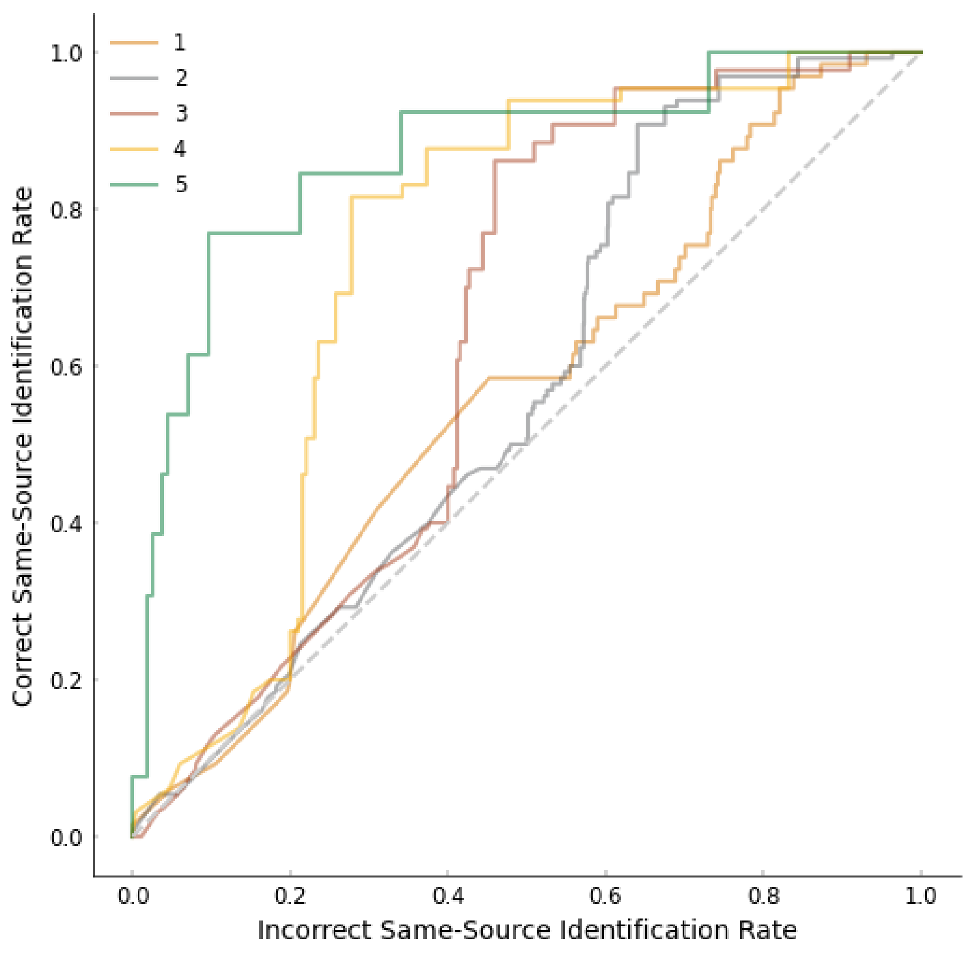

Figure 1. Receiver operating Characteristic (ROC) curve showing the proportion of times the same manufacturer scored above a varying threshold versus the proportion of the time a different manufacturer scored above that same threshold. Separate lines indicate different ROC curves for increasing number of modalities included.

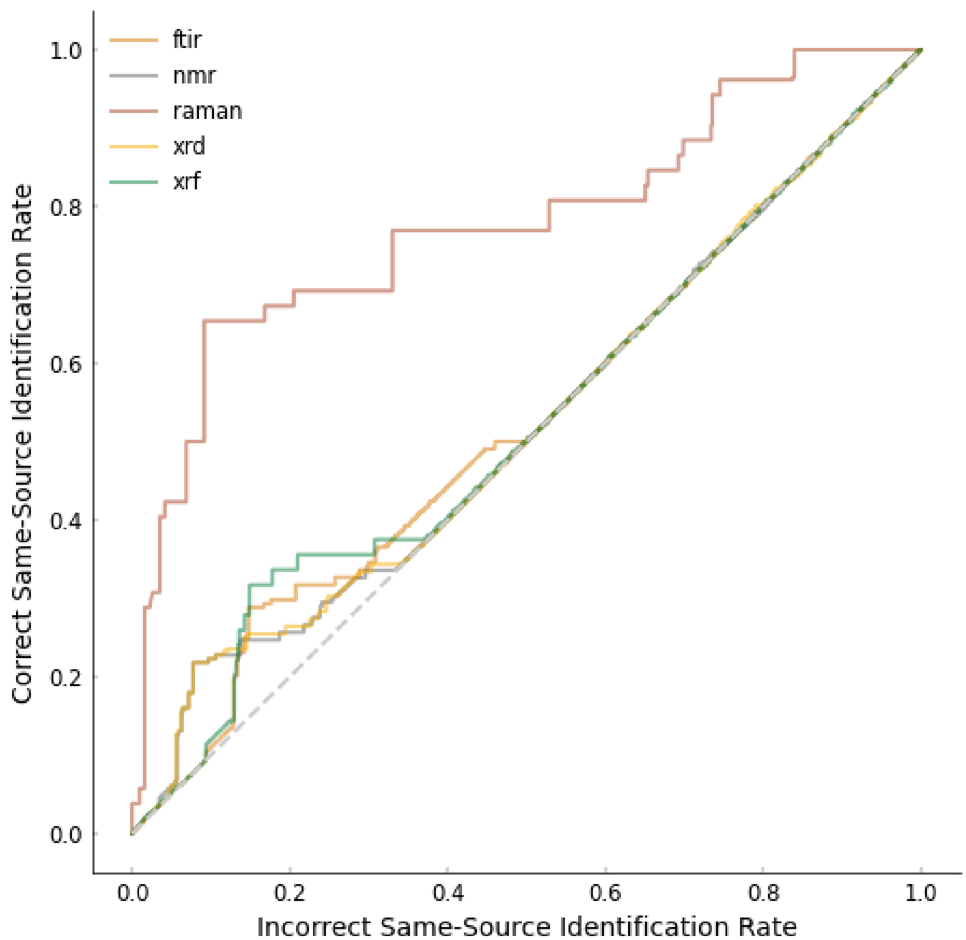

Figure 2. Receiver Operating Characteristic (ROC) curve showing the proportion of times the same manufacturer scored above a varying threshold versus the proportion of the time a different manufacturer scored above that same threshold. Separate lines indicate different ROC curves for samples including the indicated modality.

Future work should look to improve the performance of this framework. In particular, while the autoencoder framework was trained using very large data, there were only 403 combinations of modalities from the same sample for training of the sequence network. Many more cross-modal samples, whether from open literature or experiments, should be included in training to improve performance.

5.0 References

1. Chen, Ting, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. “A Simple Framework for Contrastive Learning of Visual Representations.” arXiv. https://doi.org/10.48550/arXiv.2002.05709.

2. Georgiev, Dimitar, Simon Vilms Pedersen, Ruoxiao Xie, Álvaro Fernández-Galiana, Molly M. Stevens, and Mauricio Barahona. 2024. “RamanSPy: An Open-Source Python Package for Integrative Raman Spectroscopy Data Analysis.” Analytical Chemistry 96 (21): 8492–8500. https://doi.org/10.1021/acs.analchem.4c00383.

3. Girard*, M., A. Hagen*, I. Schwerdt, M. Gaumer, L. McDonald, N. Hodas, and E. Jurrus. 2021. “Uranium Oxide Synthetic Pathway Discernment Through Unsupervised Morphological Analysis.” Journal of Nuclear Material.

4. Girdhar, Rohit, Alaaeldin El-Nouby, Zhuang Liu, Mannat Singh, Kalyan Vasudev Alwala, Armand Joulin, and Ishan Misra. 2023. “ImageBind: One Embedding Space To Bind Them All,” May. https://arxiv.org/abs/2305.05665.

5. Hollarek, Daniel, Henrik Schopmans, Jona Östreicher, Jonas Teufel, Bin Cao, Adie Alwen, Simon Schweidler, et al. 2025. “opXRD: Open Experimental Powder X-ray Diffraction Database.” arXiv. https://doi.org/10.48550/arXiv.2503.05577.

6. Johnson, Jakob, Luther McDonald, and Tolga Tasdizen. 2024. “Improving Uranium Oxide Pathway Discernment and Generalizability Using Contrastive Self-Supervised Learning.” Computational Materials Science 233 (January): 112748. https://doi.org/10.1016/j.commatsci.2023.112748.

7. Kabiri, Shayan, Nick Holden, Rory Flood, Jonathan Turner, and Sharon O’Rourke. 2024. “X-Ray Fluorescence Core Scanning Dataset and Calibration Source Code.” Zenodo. https://doi.org/10.5281/ZENODO.10956935.

8. Khan, Asifullah, Laiba Asmatullah, Anza Malik, Shahzaib Khan, and Hamna Asif. 2025. “A Survey on Self-supervised Contrastive Learning for Multimodal Text-Image Analysis.” arXiv. https://doi.org/10.48550/arXiv.2503.11101.

9. Liu, Hanxiao, Zihang Dai, David R. So, and Quoc V. Le. 2021. “Pay Attention to MLPs.” arXiv. https://doi.org/10.48550/arXiv.2105.08050.

10. Manni, Mathieu, and Nicola Viganò. 2023. “X-Ray Fluorescence Ghost Imaging - CuSn Mask - Three Wires (Fe & Cu).” Zenodo. https://doi.org/10.5281/ZENODO.7828493.

11. Moody, Kenton J., Patrick M. Grant, and Ian D. Hutcheon. 2005. Nuclear Forensic Analysis. CRC Press. https://doi.org/10.1201/9780203507803.

12. Tolstikhin, Ilya, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, et al. 2021. “MLP-Mixer: An All-MLP Architecture for Vision.” arXiv. https://doi.org/10.48550/arXiv.2105.01601.

13. Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need." In Advances in Neural Information Processing Systems, 2017-Decem:5999–6009. Neural information processing systems foundation. https://arxiv.org/abs/1706.03762.

14. Villegas Camacho, Octavio, Roberto Alejo Eleuterio, Iván Francisco Valencia, Everardo Granda Gutiérrez, Sonia Mártinez Gallegos, and Javier Illescas. 2024. "FTIR-Plastics: A Fourier Transform Infrared Spectroscopy Dataset for the Six Most Prevalent Industrial Plastic Polymers." Zenodo. https://doi.org/10.5281/ZENODO.10736650.

15. Von Konrat, Matt. 2022. "Biomonitoring of Environmental Pollution: An Exploratory Investigation Using Mosses and X-Ray Fluorescence (XRF) Spectroscopy." Zenodo. https://doi.org/10.5281/ZENODO.7126290.

16. Wishart, David S., Zinat Sayeeda, Zachary Budinski, AnChi Guo, Brian L. Lee, Mark Berjanskii, Manoj Rout, et al. 2022. "NP-MRD: The Natural Products Magnetic Resonance Database." Nucleic Acids Research 50 (D1): D665–77. https://doi.org/10.1093/nar/gkab1052.

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354
1-888-375-PNNL (7665)

www.pnnl.gov