Homology Features for X-Ray Hyperspectral Images Segmentation

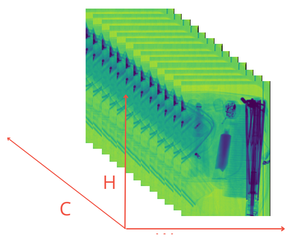
Fethi Harkat[1,2], Guillaume Gey[1], Valérie Perrier[2], Kévin Polisano[2]

[1]Detection Technology Plc, Espoo, Finland [2]Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, Grenoble, France

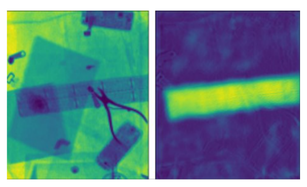
fethi.harkat@univ-grenoble-alpes.fr

1. Context and problem

Multi-energy X-ray transmission (ME-XRT) imaging uses a photon-counting detector that sorts the transmitted photons into $C = 16$ energy bands: every pixel carries a spectrum, and an image is a cube $[H, W, C]$.



ME-XRT data cube



Task: segment cigarette cartons in luggage

Why it is hard: objects **overlap along the X-ray path**, so a spectrum mixes several materials, and the bands are noisy and non-linearly correlated.

Idea: we expect spectral mixing at material interfaces to create *loops* in the spectral point cloud of a region, and we measure them with persistent homology.

2. UMAP pre-processing [1]

Each pixel of the cube $[H, W, 16]$ is a spectrum, i.e. a point of $\mathbb{R}^{16}$. **UMAP** [2] maps each point to $\mathbb{R}^5$ while keeping its neighbours, which gives a 5-channel **UMAP image** $[H, W, 5]$: the 5 values of a pixel are its **UMAP coordinates** (UMAP fitted on training pixels only).

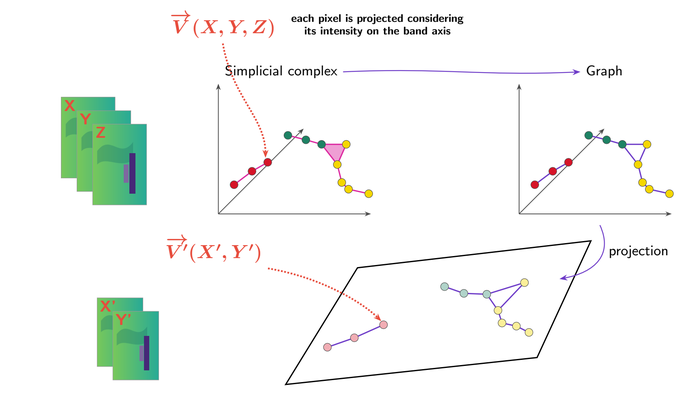
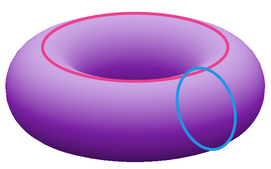


Illustration with 3 bands projected to 2: pixels with similar spectra stay neighbours.

3. Intuition: what does homology measure?

Homology counts the **holes** of a shape. A torus has two independent loops that cannot be shrunk or filled: one around the **central hole** and one around the **hole inside the tube**. On a disk, a loop can always be filled: no hole. The group H_k counts the holes of dimension k: H_0 the connected components, H_1 **the loops** (1-dimensional holes), H_2 the cavities. Its rank is the Betti number β_k. We use H_1.

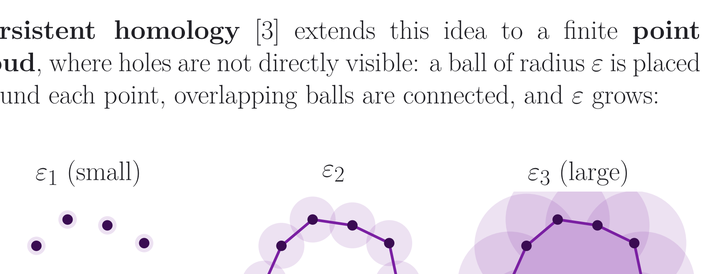


2 loops **cannot be filled** $\Rightarrow$ two H_1 classes ($\beta_1 = 2$)



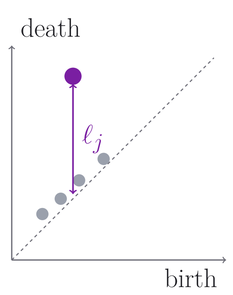
loop on a disk: **filled** $\Rightarrow$ no hole

Persistent homology [3] extends this idea to a finite **point cloud**, where holes are not directly visible: a ball of radius ε is placed around each point, overlapping balls are connected, and ε grows:



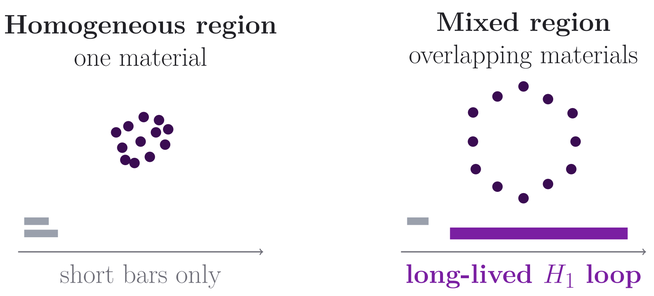
isolated points — a **loop** is born — the loop is filled: it **dies**

H_1 **persistence diagram** $\mathcal{D}_1$: each loop j is a point (b_j, d_j), with b_j the radius ε at which it is born (ε_2 above) and d_j the radius at which it dies (ε_3); its **persistence** is $\ell_j = d_j - b_j$. Far from the diagonal: persistent loop (structure); close to it: noise. Each loop can also be drawn as a bar from b_j to d_j (**barcode**).

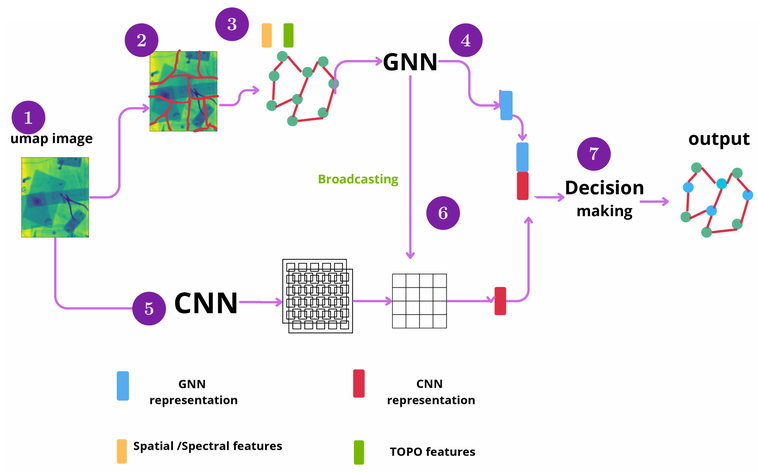


3b. H_1 separates one-material regions from mixed ones

Goal: tell apart superpixels made of **a single material** from those where **several materials overlap** (object borders, superposition). Each dot is **one pixel** of a superpixel, placed at its UMAP coordinates (shown in 2D); below each cloud, its H_1 barcode.



4. Overview: hybrid CNN–GNN with broadcasting



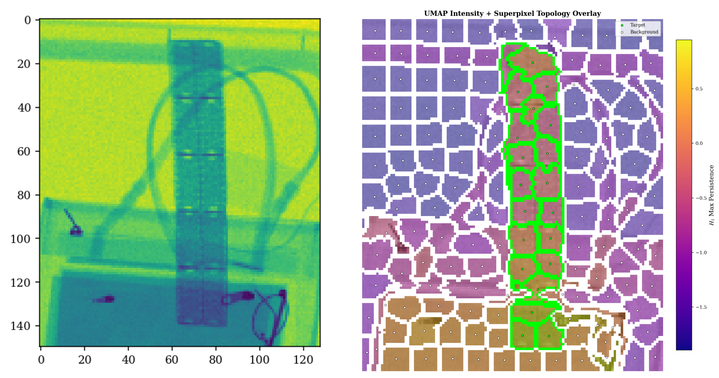
1 UMAP image (block 2); 2 **SLIC superpixels** [5], 150 per image, following object contours; 3 for each superpixel, **persistent homology** of its point cloud and spectral / geometric statistics: 41 node features (block 5); 4 **GCN** on the graph of adjacent superpixels; 5 **CNN** on the UMAP image at full resolution, trained jointly; 6 **broadcasting** of the GCN output to the pixels; 7 pixel-level decision.

5. Node features of a superpixel

For each superpixel i, the UMAP coordinates of its pixels form a point cloud $\mathcal{P}_i \subset \mathbb{R}^5$ (one point per pixel). Its H_1 persistence diagram $\mathcal{D}_1$ (block 3) is computed with **Ripser** [4]. Each node is described by **41 features** in three families:

Features	What they tell
Topological: 11 statistics of the H_1 diagram	
Max, Min, Mean, Std, Sum of ℓ_j	how strong the loops are (long = persistent, short = noise)
Mean / Max birth b_j, Mean / Max death d_j	at which scale loops appear and close: spread of the spectra
Entropy $-\sum_j p_j \log p_j$, $p_j = \ell_j / \sum_k \ell_k$	one dominant loop (low) or many similar loops (high)
Number of loops	complexity of the point cloud
Spectral: 25 statistics	
Mean, Std, quartiles Q1, Q2, Q3 (25, 50, 75 %) of each UMAP channel over the superpixel	*where* the region lies in UMAP space
Geometric: 5 descriptors	
Area, perimeter, eccentricity, solidity, orientation	shape of the region in the image

Homology describes the *shape* of the cloud, the spectral statistics say *where* it lies, i.e. which material: they are **complementary**.



UMAP image (left, sum of the 5 channels) and **MaxPers** of each superpixel (right), i.e. the persistence of its longest loop, $\max_j \ell_j$: high values (warm colours) appear in heterogeneous regions, at the borders of the objects.

4b. Broadcasting fusion

For a pixel (h, w), let $\mathcal{S}(h, w) \in \{1, \dots, N\}$ be the index of its superpixel (N superpixels in the image). The GCN gives a vector $\mathbf{z}_i \in \mathbb{R}^2$ for each superpixel i; it is copied to all its pixels:

$$\mathbf{F}_{\text{broad}}(h, w) = \mathbf{z}_{\mathcal{S}(h,w)} \in \mathbb{R}^2.$$

The CNN gives a 64-dim. feature vector $\mathbf{F}_{\text{cnn}}(h, w)$ per pixel. Both are concatenated (66 values per pixel) and a 1×1 convolution gives the scores of the 2 classes (target / background) for each pixel:

$$\bar{\mathbf{Y}}(h, w) = \mathbf{W}\,[\mathbf{F}_{\text{cnn}}(h, w);\, \mathbf{F}_{\text{broad}}(h, w)] + \mathbf{b}, \qquad \mathbf{W} \in \mathbb{R}^{2 \times 66}.$$

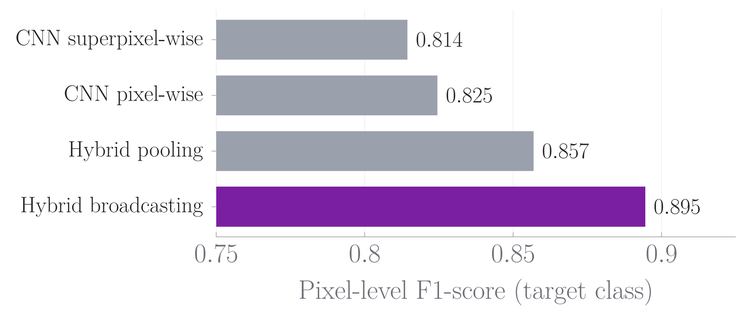
6. Feature selection: GNN branch alone

Spatial = the 30 spectral + geometric features of block 5, kept fixed; topological candidates are added (10 runs). OA: overall accuracy.

Node Features	F1 (Target)	OA
Raw HSI (16 bands)		
Topo only (30)	0.283 ± 0.011	0.391 ± 0.003
Spatial only	0.685 ± 0.005	0.912 ± 0.006
Spatial + 3 Topo	0.683 ± 0.008	0.947 ± 0.006
Spatial + 11 Topo	$\mathbf{0.764 \pm 0.007}$	0.934 ± 0.010
Spatial + 19 Topo	0.752 ± 0.003	0.948 ± 0.002
Spatial + 30 Topo	0.725 ± 0.005	0.933 ± 0.003
UMAP (5 channels)		
Spatial only	0.778 ± 0.011	0.935 ± 0.009
Spatial + 11 Topo	$\mathbf{0.823 \pm 0.003}$	$\mathbf{0.956 \pm 0.006}$

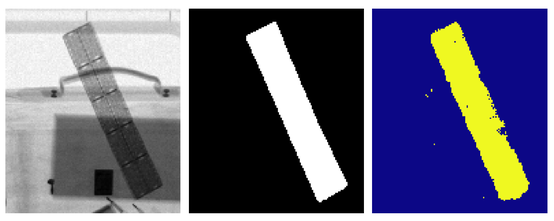
- The 11 H_1 features help **in both spaces**: the loops exist in the original spectra, they are not created by UMAP.
- More features (H_0, Betti numbers) **hurt**: they follow superpixel size, not material.

7. Hybrid model: results



Superpixel-level F1: **0.933** (broadcasting) vs. 0.872 (pooling) and 0.847 (best CNN).

+0.086 superpixel F1 and +0.070 pixel F1 over the best CNN.
Broadcasting beats pooling: it keeps the pixel resolution.



Input — Ground truth — Prediction

8. Conclusions

- H_1 persistence of superpixel spectral point clouds is a **complementary signal** to spectral-spatial features.
- A compact set of **11 H_1 statistics** is enough.
- **Broadcasting** fuses regional topology and pixel features at full resolution.

Code

References

[1] F. Harkat et al., "Manifold learning for hyperspectral images," WHISPERS 2025.
[2] L. McInnes et al., "UMAP: Uniform manifold approximation and projection," arXiv:1802.03426, 2018.
[3] H. Edelsbrunner, J. Harer, *Computational Topology: An Introduction*, AMS, 2010.
[4] U. Bauer, "Ripser: efficient computation of Vietoris–Rips persistence barcodes," J. Appl. Comput. Topol., 2021.
[5] R. Achanta et al., "SLIC superpixels compared to state-of-the-art superpixel methods," IEEE TPAMI, 2012.