



MDMC

Master in Data Management
and Curation

AREA
SCIENCE PARK

SISSA

Scanning Electron Microscopy (SEM) Images

Jacopo Zuppa

Riccardo Simonetti

Sandeep Kumar Chaluvadi

Roberta Lamberti

Label SEM images by identifying structural
similarity via a mapping algorithm



Scanning Electron Microscopy (SEM)



- To visualize any material at nanoscale, scientists use electron microscopy for imaging.
- Often, they capture lot of images with different instrumental settings, until they are satisfied with the image
- The captured images are always saved in an random folders, with random names (thinking they will remember forever).....which is not a case
- Therefore, this project comes up with an idea to help scientists working with the microscopy to ideally classify and segregate the data into an well structured formats etc.

Dataset

There are 2 datasets:

10,000 TIFF images (not categorized) in nested folders

25,000 JPEG images (categorized) in nested folders

10 different categories are labelled as:

Porous_Sponge: L0_xxx

Patterned_Surface: L1_xxx

Particles: L2_xxx

Films_Coated_Surfaces: L3_xxx

Powders: L4_xxx

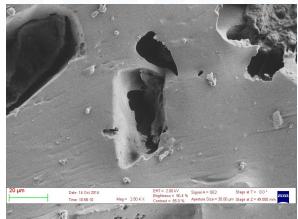
Tips: L5_xxxx

Nanowires: L6_xxx

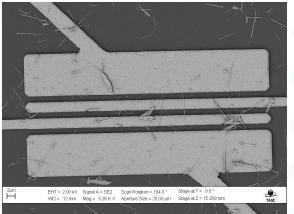
Biological: L7_xxxx

MEMS_devices: L8_xxx

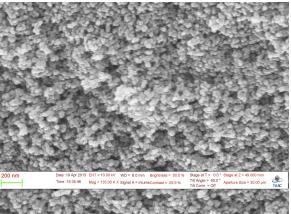
Fibers: L9_xxx



Porous_Sponge: L0_xxx



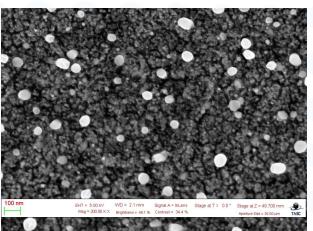
Patterned_Surface: L1_xxx



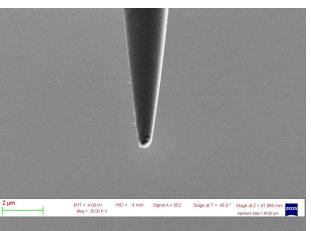
Powders: L4_xxx



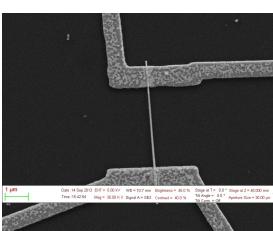
Biological: L7_xxx



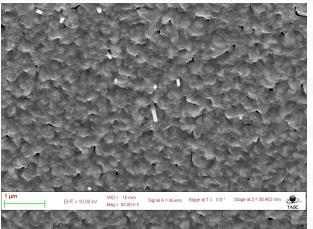
Particles: L2_xxx



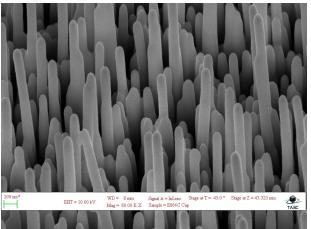
Tips: L5_xxx



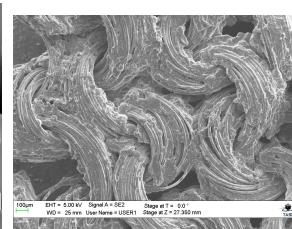
MEMS: L8_xxx



Films_coated: L3_xxx



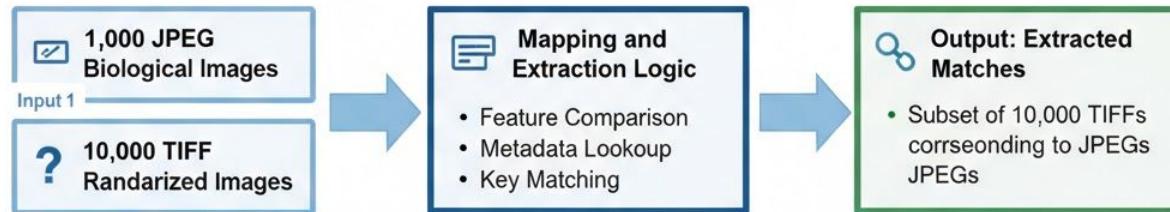
Nanowires: L6_xxx



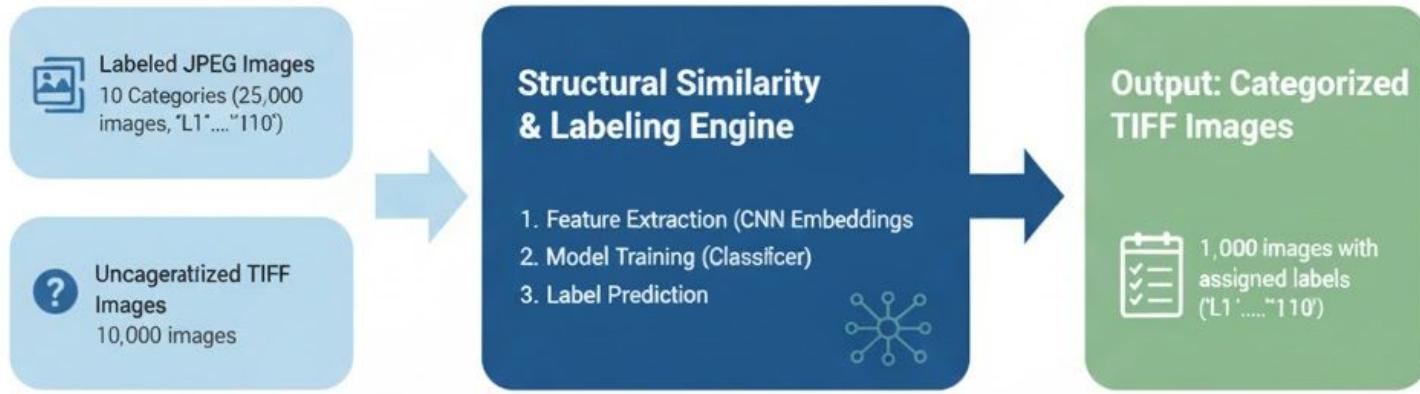
Fibers: L9_xxx

Objectives

First week: Develop an algorithm to effectively compare the 1000 Biological JPEG images with the uncategorized 10,000 TIFF images

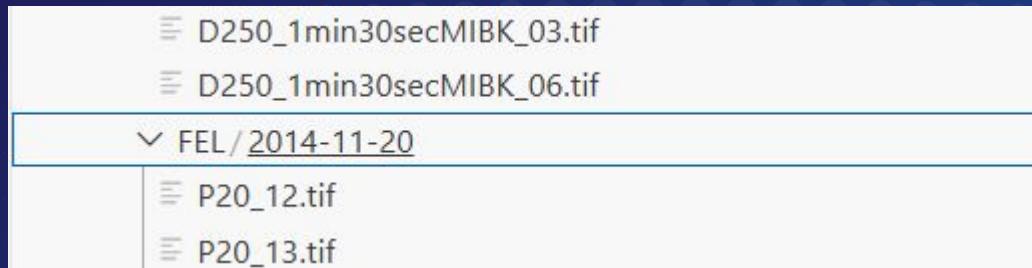


Second week: We have received 10 different categorized JPEG images (25,000) with labels ('L....'). We have to find structural similarities and assign the correct labels to the uncategorized TIFF (10,000) images



The problem

- Collection of circa 10k of unlabeled tif images



- Collection of circa 25k jpeg images labeled across 10 categories

▼ Fibres	
	L9_0a95f6c416d2bd7c7675d27602ca1b4b.jpg
	L9_00fc0a86bd4f02995acdd5b3f63401b9.jpg

The solution - pHashing and structural similarity

- In order to identify the corresponding jpeg images we use **perceptual hashing (pHashing)** and **structural similarity (SSIM)**
- **pHashing** → hash reflecting relevant visual features
- **Structural similarity (SSIM)** → brightness, contrast, structural features

The solution - Flowchart describing the processing steps

Compute hashes for tif and jpeg images



Compute Hamming distance and select the two closest jpeg images

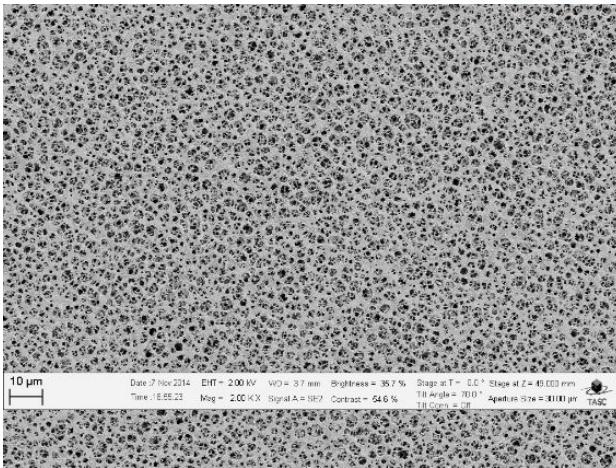


Compute SSIM values and select the highest for the best matching jpeg image



Assign jpeg image label (i.e. L0, ..., L9) to corresponding tif image

Result example



L0_388b38a8ea05fe41f3f165b9cc61a880.jpg



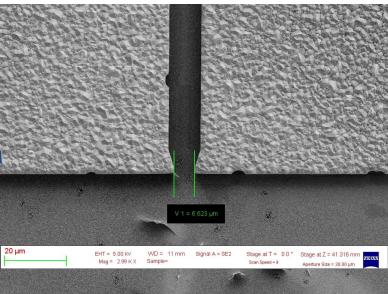
Algae+Graphene_001GP_00.tif



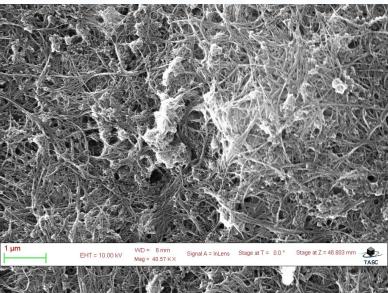
L0_Algae+Graphene_001GP_00.tif

Last implementation (still in development)

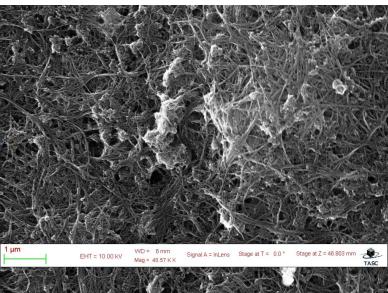
- Instead of only selecting **best match** and use it for labeling
- We select the **best three matches** and we compute **SSIM** for all (below a certain threshold of Hamming distance)
- For each match, we compute a level of certainty based on the computed SSIM
- We now have a set of **possible labels** with associated degrees of certainty



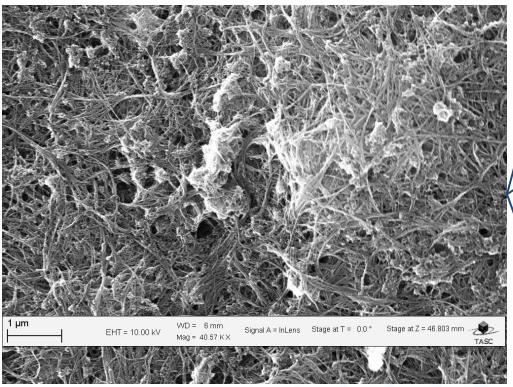
L8_8e28e0f52edca6e4a7cc560ec06b0e27.jpg;
SSIM: unknown



L6_208acd2fa4133f8ab81ca76685a38700.jpg;
SSIM: 0.9988616015284228, Very high



L6_2a4f0e72f7312d8cf1ed9eadc2b0e355.jpg;
SSIM: 0.7978146254476983, high



Dw_130207_09.tif



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



E-ARGO



beanTech®
IT moves your business

CERIC

Thank you for your attention

This Pilot training activity has been funded by the European Union – NextGenerationEU within the PNRR projects funded pursuant to Article 11, paragraph 1, of Notice 594/2024:

- “NFFA-DI cod. IR0000015, Missione 4, “Istruzione e Ricerca” – Componente 2, “Dalla ricerca all’impresa” – Linea di investimento 3.1, “Fondo per la realizzazione di un sistema integrato di infrastrutture di ricerca e innovazione” – Azione 3.1.1, “Creazione di nuove IR o potenziamento di quelle esistenti che concorrono agli obiettivi di Eccellenza Scientifica di Horizon Europe e costituzione di reti” (CUP B53C22004310006).
- “EFC cod. SSU2024-00002, Missione 4 “Istruzione e ricerca” - Componente 1, “Potenziamento dell’offerta dei servizi all’istruzione: dagli asili nido all’università” - Investimento 3.4 “Didattica e competenze universitarie avanzate” - Sub-Investimento “Rafforzamento delle scuole universitarie superiori” (CUP: G97G2400100007).