Materials Discovery: Fine-Grained Classification of X-ray Scattering Images
Hadi Kiapour1, Kevin G. Yager2, Alexander C. Berg1, Tamara L. Berg1
1 University of North Carolina at Chapel Hill, 2 Brookhaven National Laboratory

Overview
Task: Predicting a diverse set of material properties from x-ray scattering images.
Challenge: X-ray scattering data are not conventional images: no concepts of objects, location matters, intensities span many orders of magnitude.
Approach: Attribute-based visual recognition.
X-ray Scattering techniques are based on firing a beam of x-rays through a sample and deducing the structure of material based on the scattering of x-rays as a function of angle. High-performance instruments are now generating data at rates of more than a million per year.

Contributions:
- X-ray Materials Discovery Dataset: first X-ray scattering dataset fully labeled by a domain expert.
- Application of computer vision to scientific datasets (abstract imagery).
- Automatic system for analyzing, searching and classifying new unlabeled x-ray scattering images at large scale.

Dataset
Images: 2832 x-ray scattering images from 13 measurement runs.
- A run is a set of related x-ray scattering images collected for closely-related material samples, continuously captured over a short period of time.
Attributes:
- 98 binary attributes annotated by a domain expert for each image.
- Attributes indicate type of measurement, scattering appearance features, chemical and physical properties of the materials.

Models
Two-level classification: SAXS, WAXS.
Approach: SVM, LBPPHOG, SIFT.

Model Selection: Cross-validation across different runs instead of randomly partitioning the samples to avoid biasing the model.
Visual Descriptors: Down-sampled, HOG, PHOG, GIST, LBPPHOG, SIFT.

Classification
Mean Average Precision over all attributes for two-level vs. no hierarchy classification:

Automated Annotation
Predicted Tags:
- Beam off image
- Diffuse low-q. iso.
- Halo
- Higher orders: 10 to 20
- Higher orders: 4 to 6
- Linear beamstop
- Ring: Isotropic

Robust retrieval allows new data to be compared to database of previous experiments.

Automatic Tagging
- AgBH, Beam off image, Higher orders: 10 to 20, Ordered, Photonic CCD, Ring: Isotropic, TWAXS.

Notes
- All the retrieved samples have the ring near the center (i.e. diffuse low-q, isotropic).
- Retrieved images only have sharp rings and are of the exact same material as the query.
- Retrieved images are not of gratings, they have bright scattering intensity near the origin, but lack the bright stripes. The input and the retrievals do indeed have lots of rings which are called higher orders.
- Retrieved examples must have isotropic halo (very broad rings) but without sharp uniform rings.

This reliable behavior enables automated, high-speed data analysis.