Materials Discovery: Fine-Grained Classification of X-ray Scattering Images

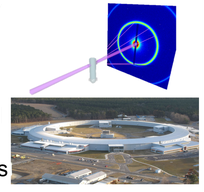
Hadi Kiapour[1], Kevin G. Yager[2], Alexander C. Berg[1], Tamara L. Berg[1]
[1] University of North Carolina at Chapel Hill, [2]Brookhaven National Laboratory

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

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.

SAXS: Small Angle X-ray Scattering

2D detector obstruction
AgBH
Higher orders: 4 to 6
Ring: Isotropic

2D detector obstruction
Diffuse specular rod
Linear beamstop
Thin Film

Bragg rods
Grating
Linear beamstop
Peaks: Along ring

WAXS: Wide Angle X-ray Scattering

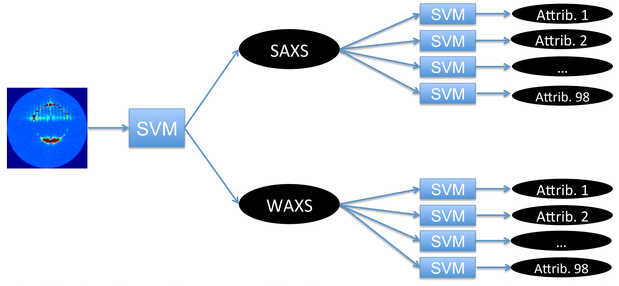
Beam off image
Halo: Isotropic
Higher orders: 2 to 3
P3HT

AgBH
Beam off image
Higher orders: 10 to 20
Ring: Isotropic

Beam off image
Halo: Isotropic
P3HT
Ring: Isotropic

Approach

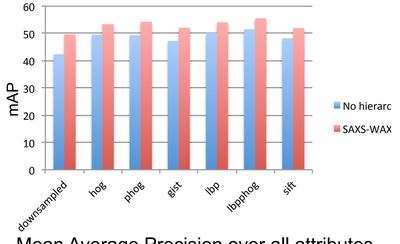
Two level classification:



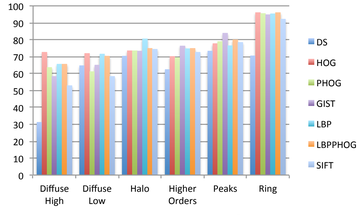
Model Selection: Cross-validation across different runs instead of randomly partitioning the samples to avoid biasing the model.

Visual Descriptors: Down-sampled, HOG, PHOG, GIST, LBPPHOG, SIFT

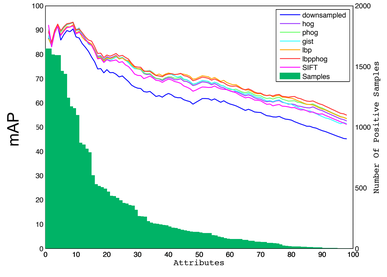
Classification



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



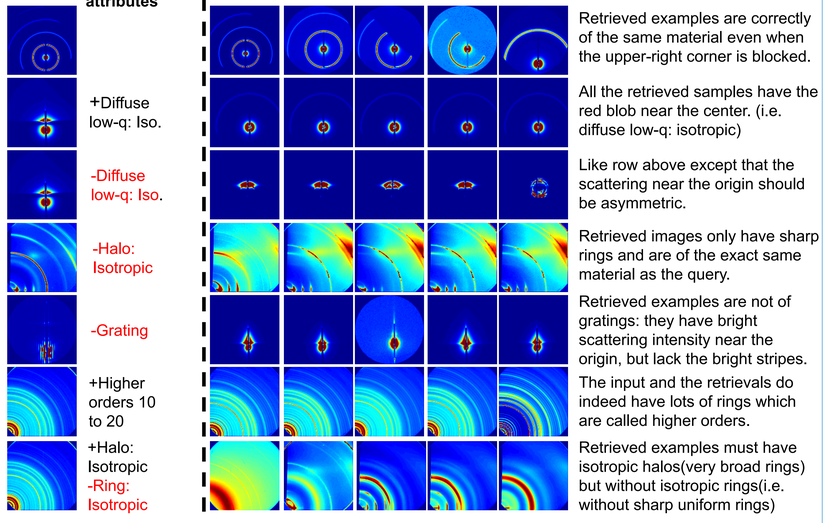
Average precisions for classification of high-level attributes



Two level classification:
Mean average precision for attribute classification. Attributes are ordered by the number of positive samples.

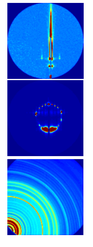
Our method can classify (>80% precision) into many scientifically-meaningful categories.

Retrieval



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

Automatic Annotation



Predicted Tags

Block-copolymer, GISAXS, Linear beamstop, MarCCD, Ordered, Specular rod,Thin Film, Vertical streaks, Weak scattering, Bragg rods, Silicon

GISAXS, Grating, Higher orders: 4 to 6, Linear beamstop, MarCCD Peaks: Along ring, Thin film, Vertical streaks

AgBH, Beam off image, Higher orders: 10 to 20, Ordered, Photonics CCD, Ring: Isotropic, TWAXS

Automated tagging can identify attributes the human expert missed. This reliable behavior enables automated, high-speed data analysis.