Automated data analysis of scanning electron microscopy images on airborne magnetic particulate matter

Ahmed Musa[1,2], Baeckkyoung Sung[1,3*], Leon Abelmann[1,4*]

[1]KIST Europe and [2]Saarland University, Saarbrücken, Germany; [3]University of Science & Technology, Daejeon, South Korea;[4] Delft University of Technology, Delft, The Netherlands

ahmed.musa@kist-europe.de

Introduction

- In our research, we investigate how to reliably count the no. of particles, or more generally, how to analyse SEM images reliably.
- Particulate Matter are tiny particles in the air, typically less than 10 micrometer (PM10).
- One typical imaging technique is Scanning Electron Microscopy, which uses an electron beam and measures emitted back-scattered and secondary electrons.
- To analyse samples collected from the air, current automated image analysis tools are highly variable and user dependent.

Figure 1: Filters deployed with and without magnets [3]

Iterative sampling for ground truth count data

We study confidence intervals of our estimated counts given a sample of several people counting, as well as the number of people required to reliably annotate our data.

Confidence Interval of count

If we assume our data generating distribution is normal we can easily estimate the confidence interval using the standard error of the mean equation:

$$Particlecount = \mu_s \pm t * \frac{\sigma}{\sqrt{n}} \tag{1}$$

This equation allows to use our sample to calculate a confidence interval that likely contains the true particle count.

Repeatability Analysis of Segmentation Methods

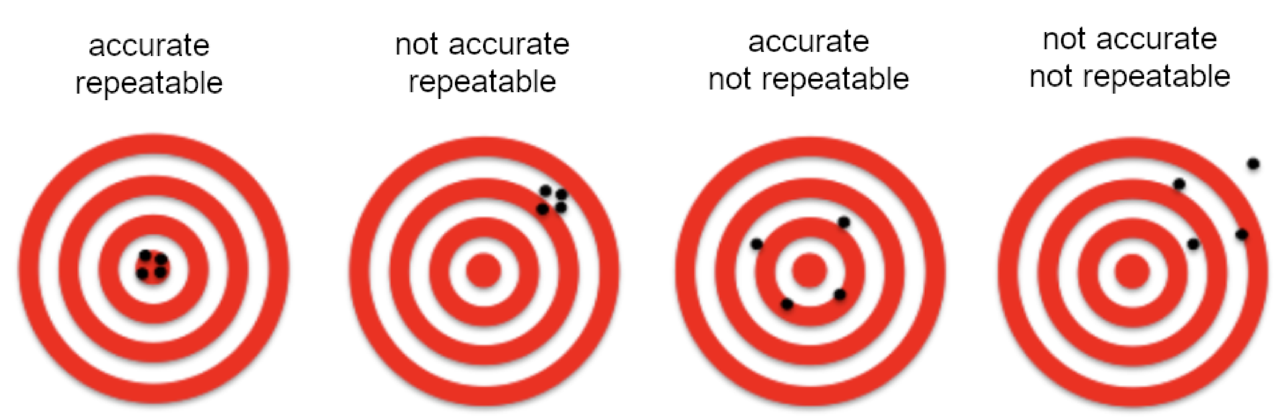

Figure 2: Accuracy vs Repeatability [3]

To analyse repeatability we consider variation in input representation, by image transformations that does not corrupt or change the input with respect to how we perceive it such as interpolating. Our intuition is as follows: **a good segmentation method should be able to detect the same particles even if the image is transformed as long as the transformation preserves the particles.**

Image Transformations

The transformations we consider are as follows:

1. Rotating clockwise, counterclockwise, flipping, and mirroring has no effect on the results so we drop them from our table.
2. **cutandmix**: we cut the image from the middle and switch right and left halves.
3. **stretchx**: doubling image width and interpolating extra pixels.
4. **stretchy**: doubling image height and interpolating extra pixels.
5. **stretchxy**: doubling image width and height and interpolating extra pixels.
6. **subdivide**: cut image to four quarters, count subparts and add up the counts.

Segmentation Algorithms

Figure 3: Segmentation flow

In prepossessing we apply contrast limited adaptive histogram equalization (CLAHE) to improve contrast. We also apply erosions three times. [5]

In thresholding we use a method to **find a threshold value to classify each pixel as foreground or background** depending on if the intensity is greater than the threshold.

In postprocessing we apply morphological opertations, removing small holes, objects.

We then label **every connected component as a particle, we consider pixels to be connected if they are vertically or horizontally connected.**.

Results

We test segmentation methods on many images and calculate **sensitivity to transformation** using max(ratio,1) - min(ratio,1), where the ratio is the no. of particles by the segmentation after / before transformation. **The larger this value the worse the repeatability of the method**. Multi Otsu and li seem to perform the best as they show relatively low sensitivity to all the image transformations.

method	cutandmix	stretchx	stretchxy	stretchy	subdivide
IsoData	3.32	0.70	1.44	1.25	1.00
li	0.34	1.55	1.75	1.54	1.55
Mean	0.14	0.69	0.83	0.70	4.61
Multi Otsu	1.49	1.30	1.07	0.58	1.62
Niblack	0.31	6.00	37.10	7.15	0.87
Otsu	1.09	1.44	6.41	1.75	2.07
Triangle	25.44	3.63	4.03	2.84	5.84
Yen	0.62	4.68	8.63	4.30	0.99

Table 1: Repeatability Analysis Results: median sensitivity to transformations

Protocol for data collection for Accuracy evaluation

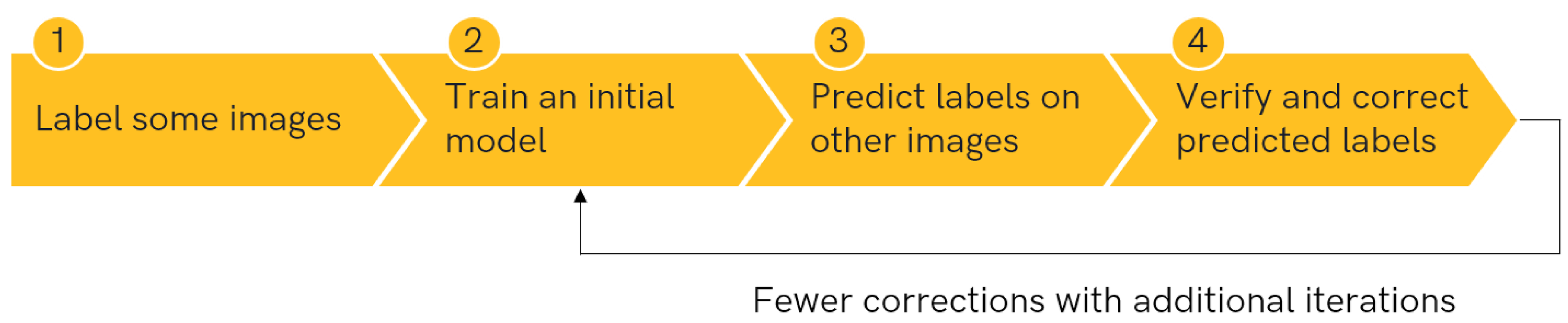

Figure 4: Model assisted labelling[1]

Conclusions

- We can use iterative labelling to count particles until we have acceptable uncertainty.
- We can apply repeatability analysis for deterministic segmentation algorithms by changing input images representation using transformations that preserves particles.
- Multi Otsu and li showed least sensetivity to image transformations.

References

[1] Model assisted labelling, segmentai. https://docs.segments.ai/tutorials/model-assisted-labeling, note = Accessed: 25-11-2022.

[2] Skimage.filters documentation. https://scikit-image.org/docs/stable/api/skimage.filters.html. Accessed: 24-11-2022.

[3] Understanding precision: accuracy, repeatability, precision.

[4] Wikipedia, confidence interval. https://en.wikipedia.org/wiki/Confidence$_interval, note = Accessed : 24 - 11 - 2022$.

[5] Han J. Han T. Y. J. Kim, H. Machine vision-driven automatic recognition of particle size and morphology in sem images. *Nanoscale*, 12(37):19461–19469, 2020.

[6] et al. Mehta, Nihaal. Repeatability of binarization thresholding methods for optical coherence tomography angiography image quantification. *Scientific reports*, 10(1):1–11, 2020.