Multidimensional particle characterisation from *in-situ* imaging using deep learning and transfer learning.

Christopher Boyle[*,1,2], Carla Ferrera[1,2], Yi-Chieh Chen[2], Christos Tachtatzis[3], Ivan Andonovic[3], Cameron Brown[1,4], Jan Sefcik[1,2], and Javier Cardona[1,2,3]

*christopher.boyle@strath.ac.uk
1. Future Manufacturing Research Hub in Continuous Manufacturing and Advanced Crystallisation (CMAC), University of Strathclyde, UK
2. Department of Chemical and Process Engineering, University of Strathclyde, UK
3. Department of Electronic and Electrical Engineering, University of Strathclyde, UK
4. Strathclyde Institute of Pharmacy and Biomedical Sciences (SIPBS), University of Strathclyde, UK

1. Motivation

Particle size and shape are important in the pharmaceutical industry, affecting both process efficiency and product performance.

Quality-by-design and continuous manufacturing are aided with appropriate models of processes — selection and calibration of which are informed by measurement of particle size and shape.

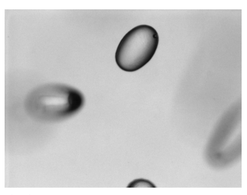
Off-line measurements have inherent limitations when following the trajectory of particle attributes in a process; removing and treating material for off-line analysis can alter particle characteristics.

In contrast, in-line measurements provide representative measures of particle size and shape at the expense of producing more challenging (out of focus, overlapping particles) datasets for extraction of particle characteristics.

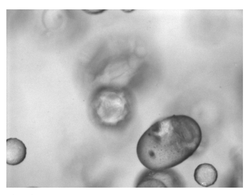
2. In-line Images

An in-line probe (PVM, Mettler Toledo) was used to record the images used here. Two materials were investigated: a mixture of polystyrene spheres and ellipsoids, and lactose particles. A subset of the collected images were used to train a machine learning model, and applied to the rest to obtain particle size and shape.

Training datasets were formed of PVM images and segmentation masks. To obtain segmentation masks of these images, two methods were employed: (1) manual annotation, (2) automatic annotation using an classical (i.e. non-ML) image analysis algorithm [1]. However the former method results in a much improved model (due to the high quality annotations) and so results using this training set are shown here.

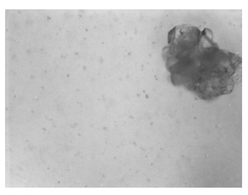


(a) 125–180 μm at 1 wt%

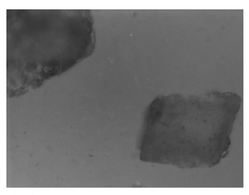


(b) 250–355 μm at 10 wt%

Figure 1: Polystyrene particles



(a) 300–355 μm at 5 wt%



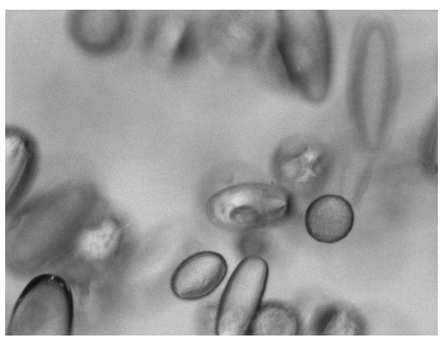
(b) 400–500 μm at 7.5 wt%

Figure 2: Lactose particles

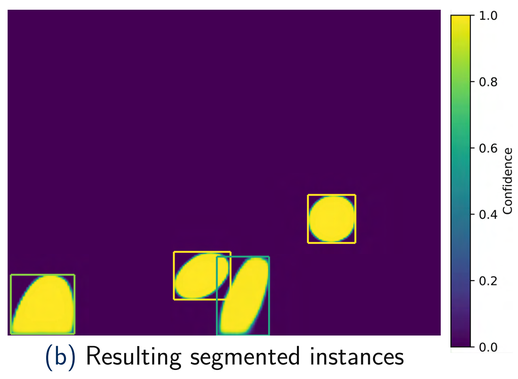
3. Machine Learning Model

The machine learning model used is "Detectron 2" [2], a deep learning model in the RCNN. The model is trained in two configurations: either from scratch or pre-trained on the Common Objects in COntext (COCO) dataset (leveraging transfer learning). Example segmentation results shown below.

Figure 3: Examples of Detectron 2 applied to the COCO dataset. From [2].

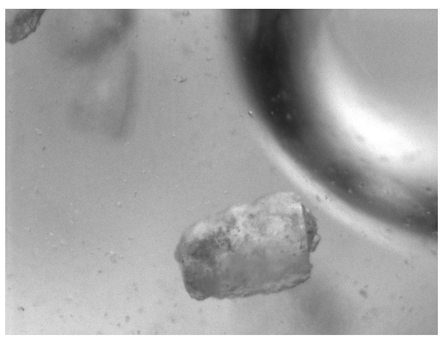


(a) Input PVM image

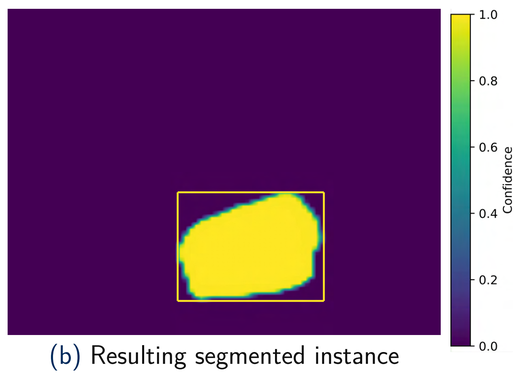


(b) Resulting segmented instances

Figure 4: Segmentation of polystyrene particles, 125-180μm at 5wt%. All in-focus particles are segmented, while the out-of-focus particles are rejected. The ellipsoid in the bottom-middle has a lower overall confidence score (indicated by bounding box colour), likely due to the two small out-of-focus spheres encroaching on it.



(a) Input PVM image



(b) Resulting segmented instance

Figure 5: Segmentation of a lactose particle, 250-300μm at 20wt%. The particle in-focus is segmented correctly, while the large bubble is left unsegmented.

The segmentation masks in the figures above are colour coded: bounding box colour indicates overall confidence in the instance prediction while the mask colour indicates pixel level confidence in the segmentation.

4. Polystyrene Particle Size Distributions

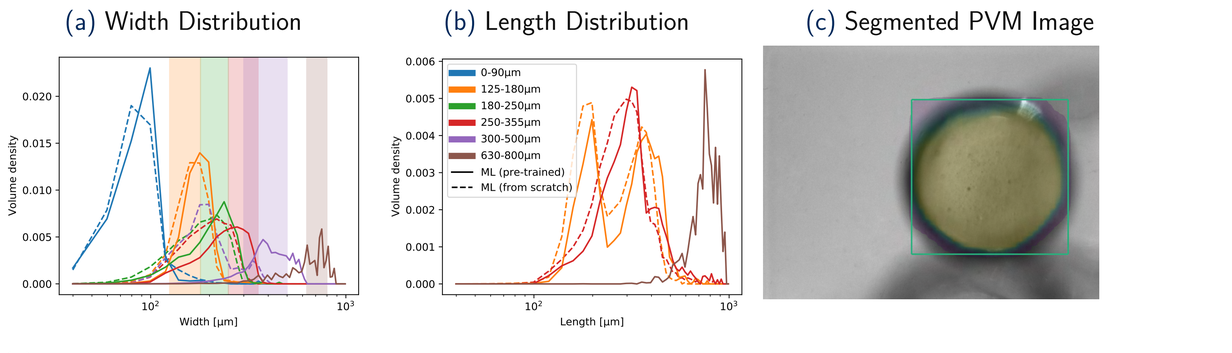


Figure 6: (a) Width, and (b) length distributions for polystyrene spheres and ellipsoid particles obtained using Detectron 2 from in-line PVM. Two training methods are compared: Detectron 2 trained from scratch (dashed lines) and pre-trained on COCO (solid lines). Shaded area shows sieved size range. A reduced set of sizes are shown in (b) for clarity. (c) shows an example of the pre-trained model segmenting a large particle (630-800μm), where the model trained from scratch fails to segment any particle at all.

Training from scratch results in a model only capable of segmenting particles similar to the training dataset and it fails at larger sizes — mitigated by use of the pre-trained model which has no problem at larger sizes (see (c)). The width distribution (a) shows sizes closely matching the sieved ranges, albeit with some shouldering outside the range (emphasised by the volume weighting). The length distribution shows a bimodality for the 125-180μm as expected for the mix of spheres (length in sieve range) and ellipsoids (with length above range). This bimodality reduces as size increases, which might make sense as large sized samples are predominantly spherical, however this is at a smaller size than expected as revealed below.

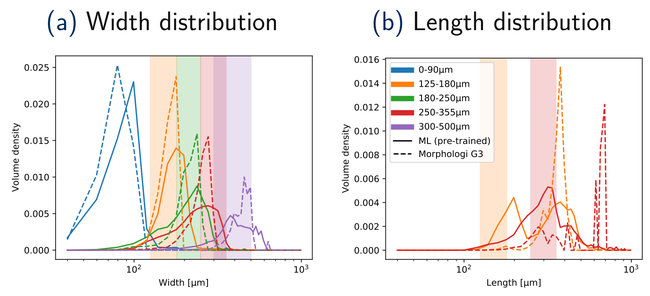


Figure 7: (a) Width, and (b) length distributions (volume weighted) for polystyrene spheres and ellipsoid particles obtained using pre-trained Detectron 2 from in-line PVM (solid lines) and by Morphologi G3 off-line microscopy (dashed lines). Colour indicates size range of particles as in legend. A reduced set of sizes are shown in (b) for clarity.

Left: comparing the performance of the Detectron 2 to off-line imaging (Morphologi G3) to validate the method. Off-line analysis gives a good representation of the system (as particles are laid out flat and far apart — easy to segment). In-line imaging does not have any of these advantages yet manages to obtain many of the same features as the G3. The width distributions are very similar. Length shows fewer ellipsoids picked up by Detectron than G3 at large sizes — likely due to the focus on smaller particle sizes in the training set.

5. Lactose Particle Size Distributions

The results of the trained (pre-trained on COCO, fine-tuned on polystyrene) Detectron 2 are shown on the figure to the right. The width distribution (a) matches with the sieves, albeit with some shouldering (emphasised by the volume weighting). The length distribution (b) shows an apparent elongation of the smaller particles with peaks to the right of the sieved size range. For larger lactose particles, this effect is not seen and the particles have an apparent aspect ratio closer to one.

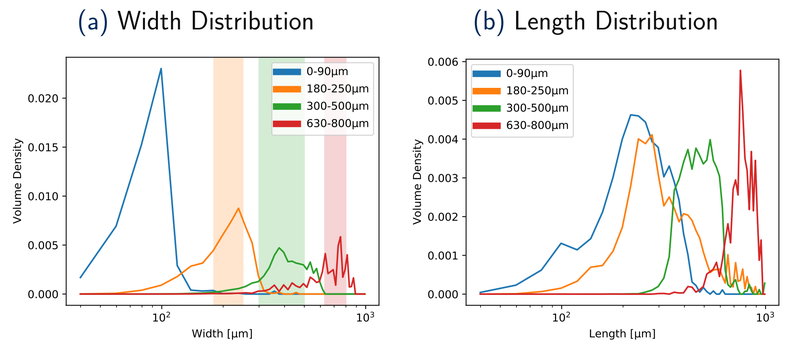


Figure 8: (a) Width, (b) length, distributions (volume-weighted) for lactose particles obtained using Detectron 2 (pre-trained) from in-line PVM. Colour indicates sieved size range of particles as in legend.

6. Conclusions

On this poster, a method for extracting particle size information from in-line imaging measurements was presented. The results were validated by comparison to off-line microscopy, showing that in-line image analysis in this way can yield quite representative size and shape measurements. It has also been shown that even a very small dataset of high-quality annotated images yields good results, especially when leveraging the power of other larger datasets with transfer learning.

Acknowledgements

This work was funded jointly by GlaxoSmithKline, Bayer, AstraZeneca, Novartis, Eli Lilly, Takeda, Pfizer, and Roche and was carried out within the CMAC Future Manufacturing Research Hub (Grant ref: EP/P006965/1). We are grateful to Francesca Perciballi (AstraZeneca), Neda Nazemifard (Takeda), Vaclav Svoboda (Pfizer), Chris Burcham (Eli Lilly), Kevin Girard (Pfizer), David Lauri Pla (Pfizer), Alexandre Vieulfaure (Pfizer) and Ben Renner (Takeda) for project guidance.

References

[1] Javier Cardona, Carla Ferreira, John McGinty, Andrew Hamilton, Okpeafoh S Agimelen, Alison Cleary, Robert Atkinson, Craig Michie, Stephen Marshall, Yi-Chieh Chen, et al.
Image analysis framework with focus evaluation for in situ characterisation of particle size and shape attributes.
Chemical Engineering Science, 191:208–231, 2018.

[2] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick.
Detectron2.
`https://github.com/facebookresearch/detectron2`, 2019.