FLANDERS MAKE

DRIVING INNOVATION IN MANUFACTURING

Active learning for quality inspecting with synthetic hot-start approach

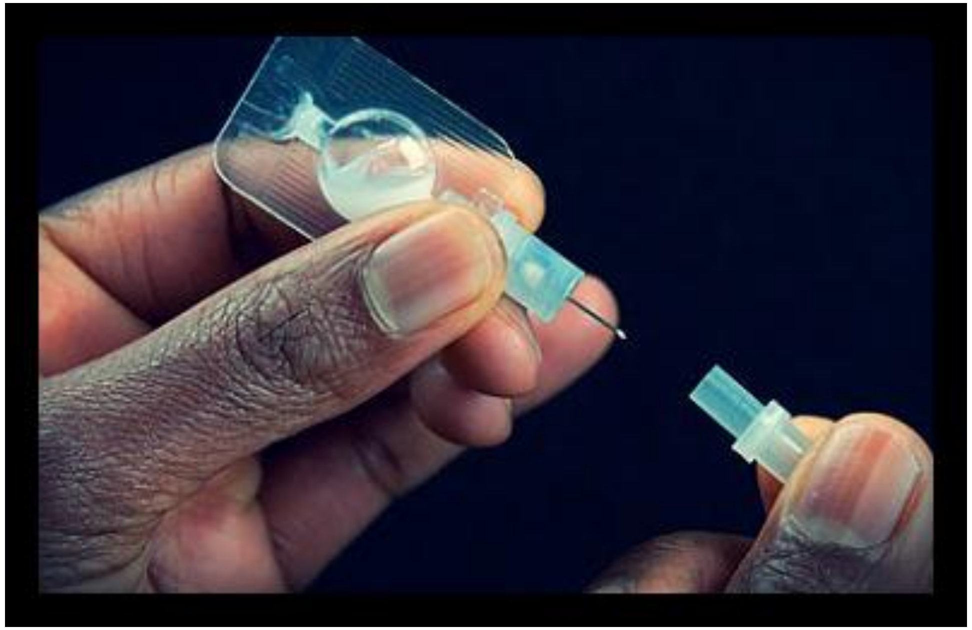
A. De Rybel, S. Moonen
N. Michiels, S. Dehaeck

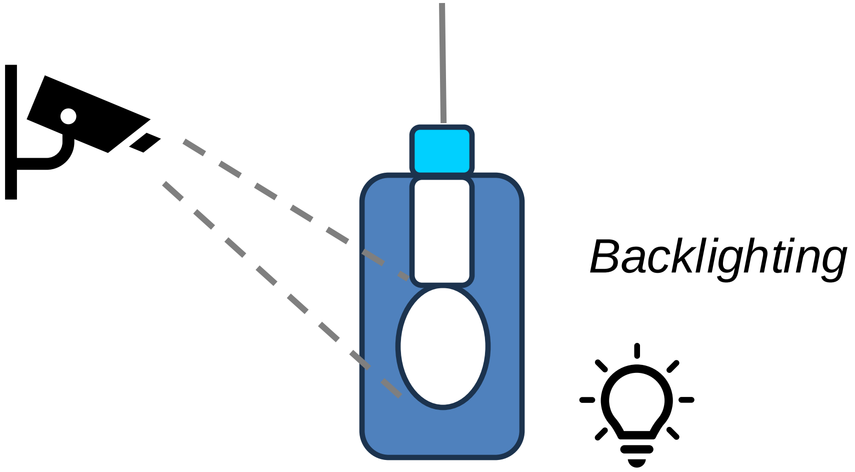
Table of contents

Introduction to the application

Unijects single-use syringes

- Production is (mostly) automated
- Quality inspection automation using computer vision
- **Inclusions,** Contaminants inside liquid bulb container





Real and synthetic data generation

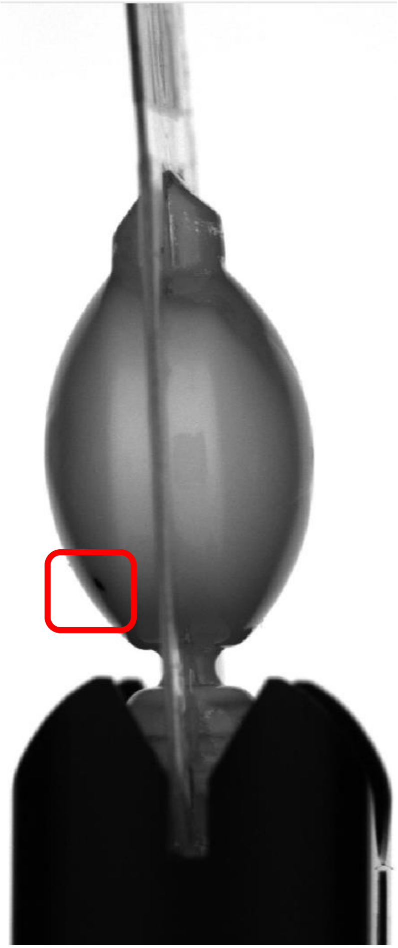
Real dataset creation 1640 samples, manually trigger defects.

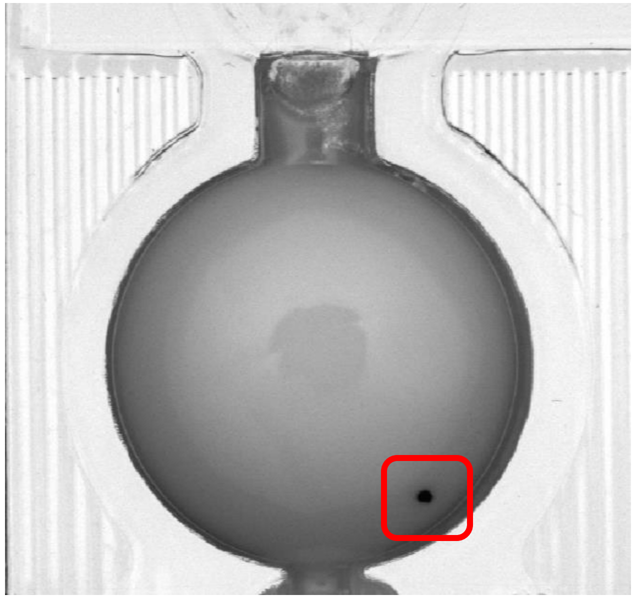
- OK samples
- Inclusions
- Outside contaminations
- 4 views (sides, front and back)

For a data scientist this is a luxury!

- Defects in production lines, especially pharmaceutical are very rare
- Manually triggering defects means halting production

MAIN research question: Are synthetic generated defects useful, knowing limited information about real defect shapes?





Real and synthetic data generation

Difference Render method
Synthetic data generation

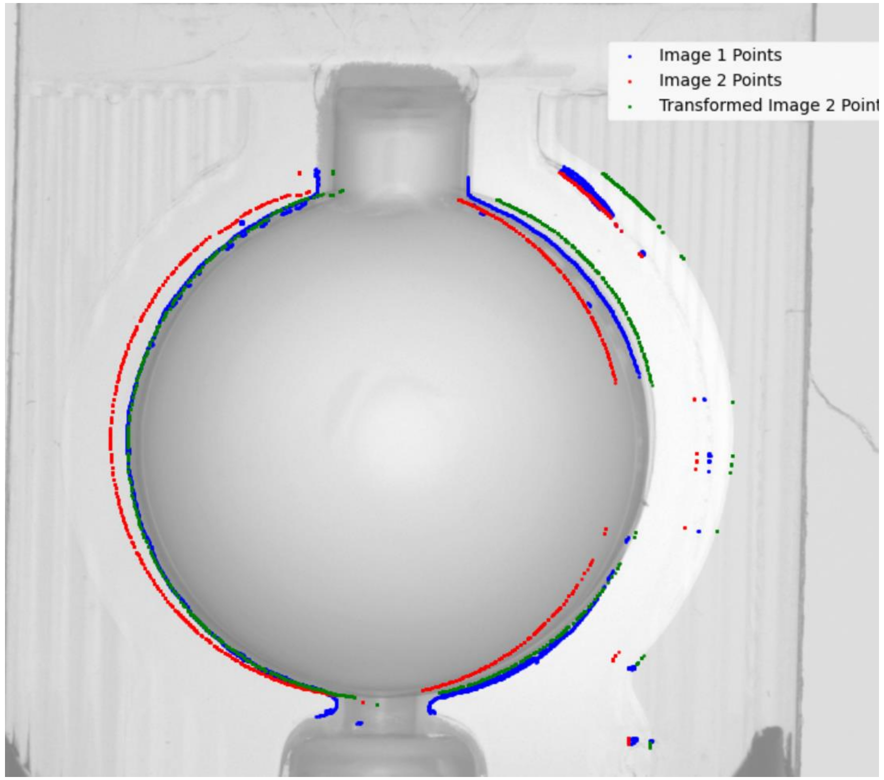
Step 1.
Create Uniject scene in Unity with backlight

Step 2.
Render defects on bulb in Unity. Ray tracing makes this realistic

Real and synthetic data generation

Difference Render method

Synthetic data generation



Step 3.
Allign Unity image
With real image OK
sample using
contour of bulb.

64 unique real images used

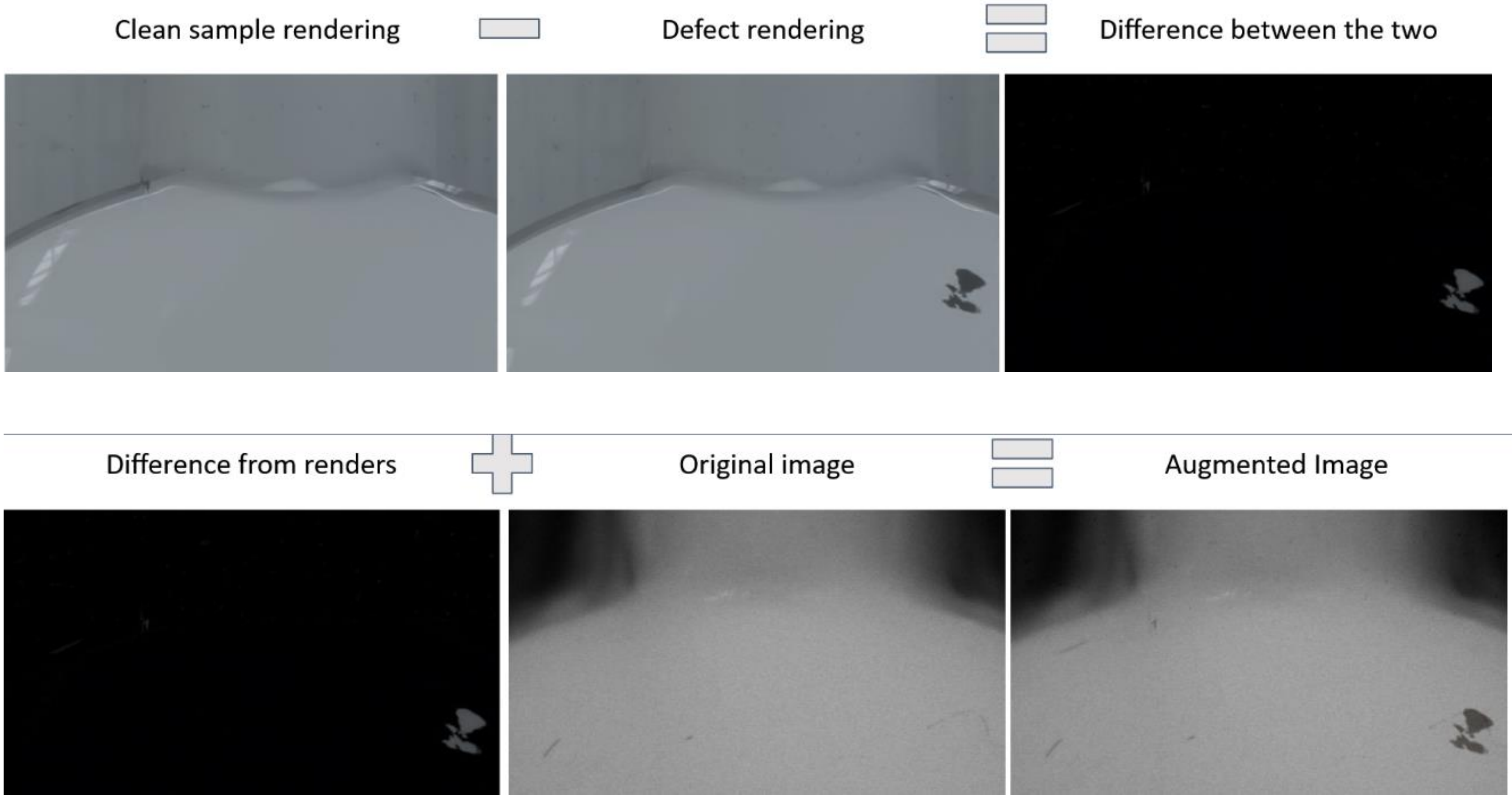
For each image you
can generate
multiple defect
variations

Real and synthetic data generation

Difference Render method

<u>Step 4.</u>
Take difference of unity image and rendering and add it to real image

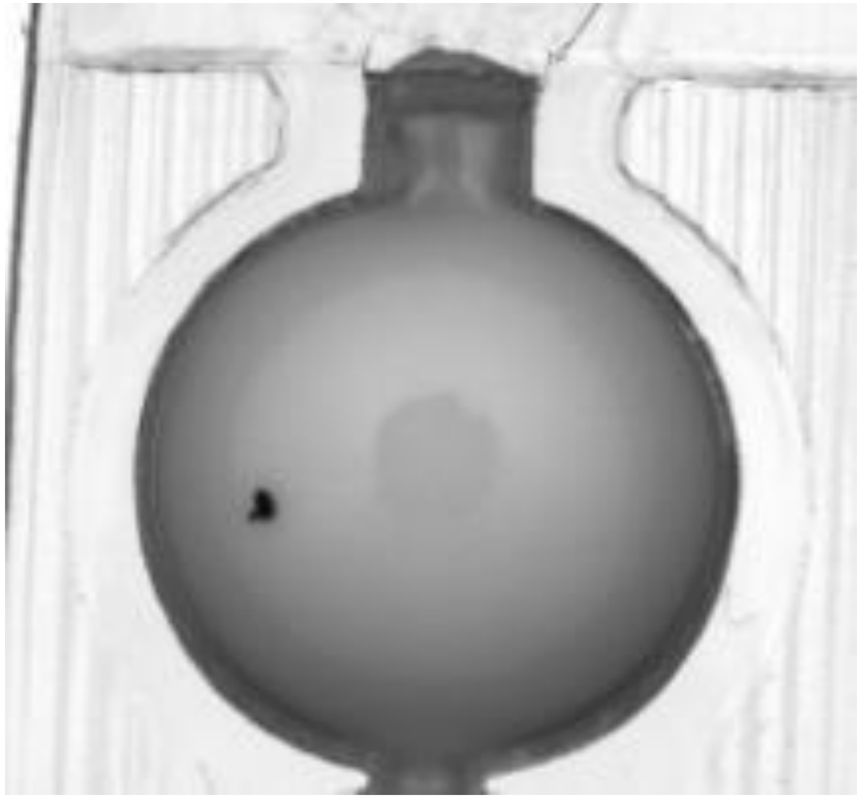
Result are real images with rendered defect on top.



Real and synthetic data generation

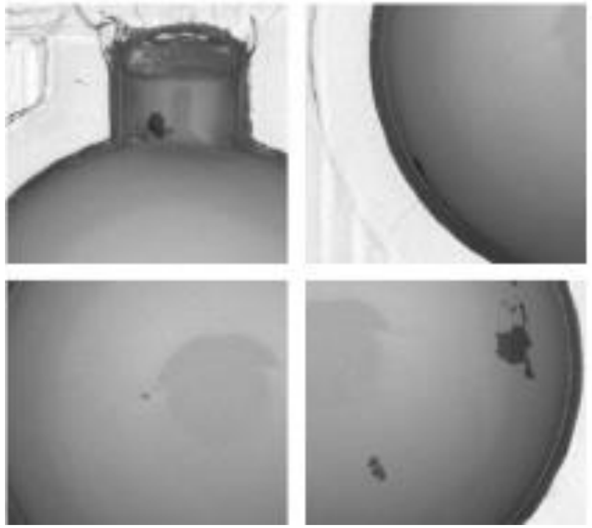
Difference Render method. **spot like defects**

Real



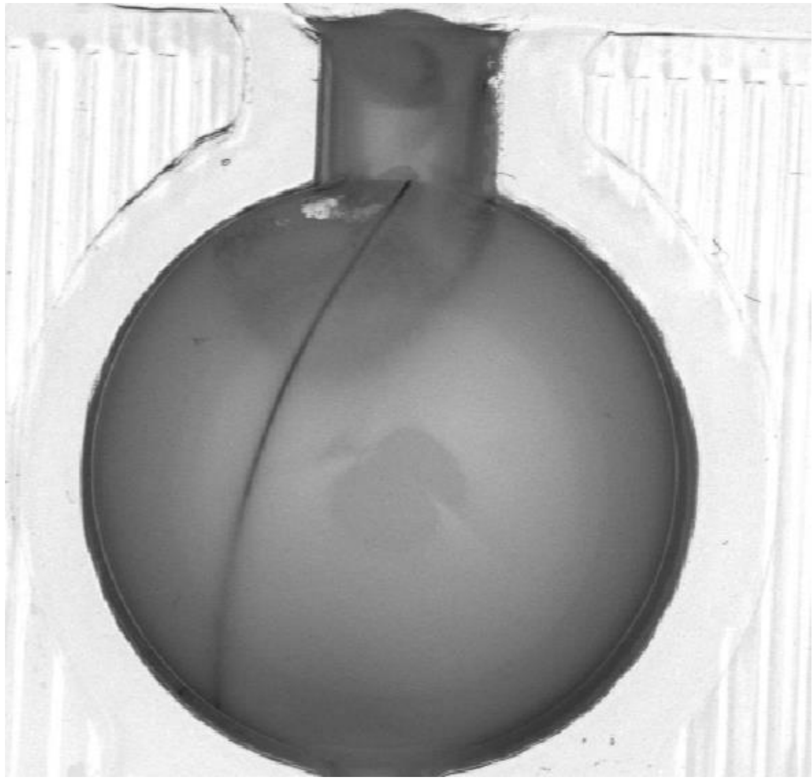
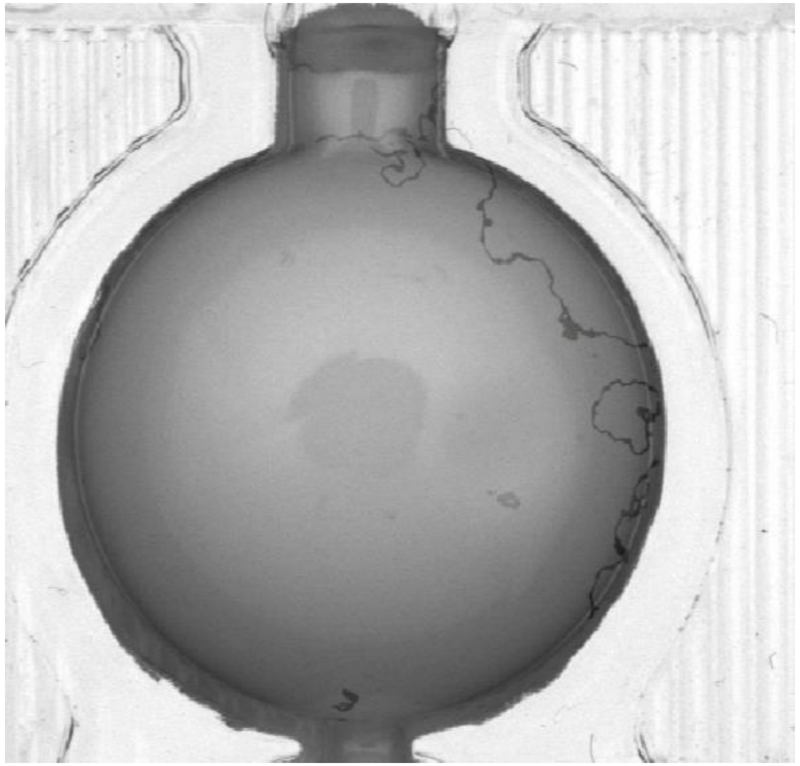
Rendered

Rendered



Real and synthetic data generation

Difference Render method. <u>line like defects</u>

Real	Rendered	Rendered
		

Introduction to active learning sampling

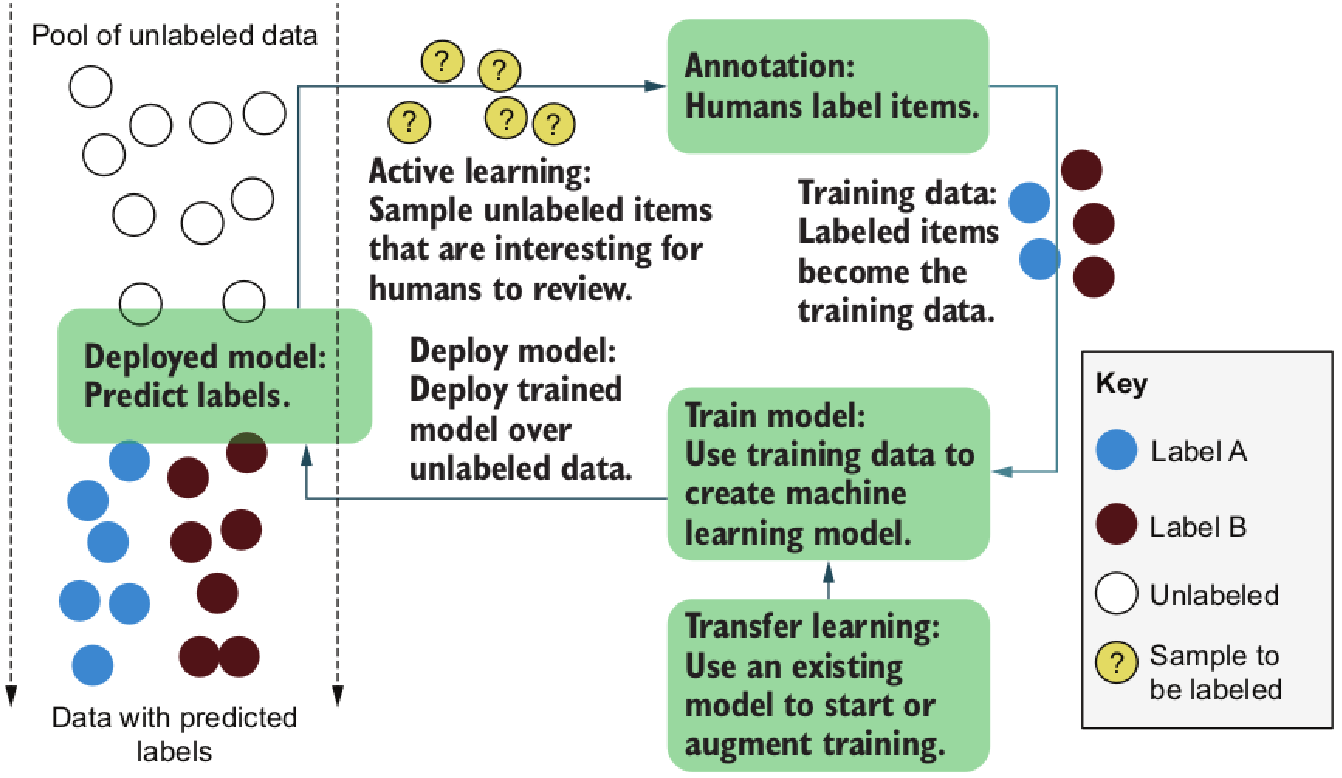
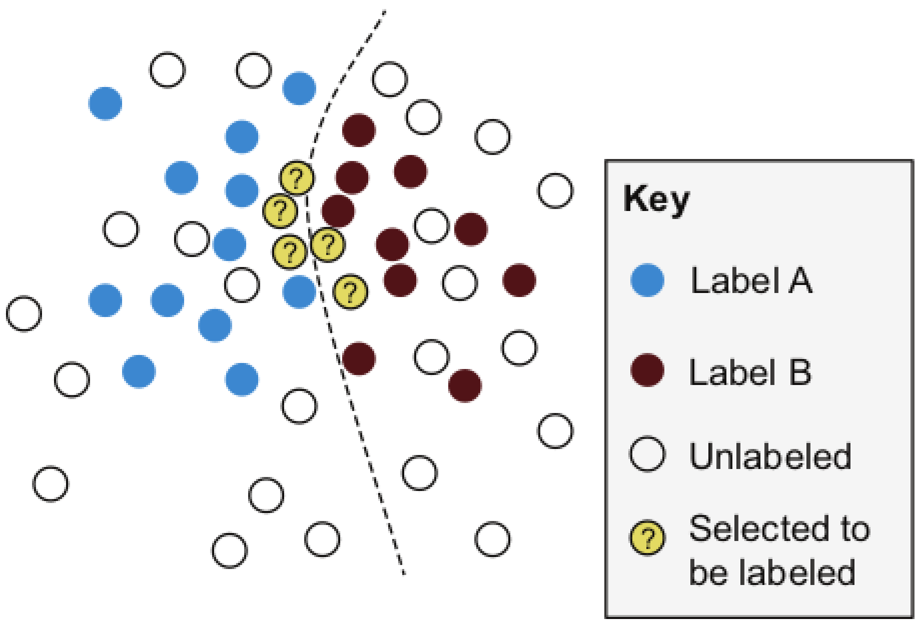


Figure 1.1 A mental model of the human-in-the-loop process for predicting labels on data



**source: Human in the loop machine learning, Robert M.*

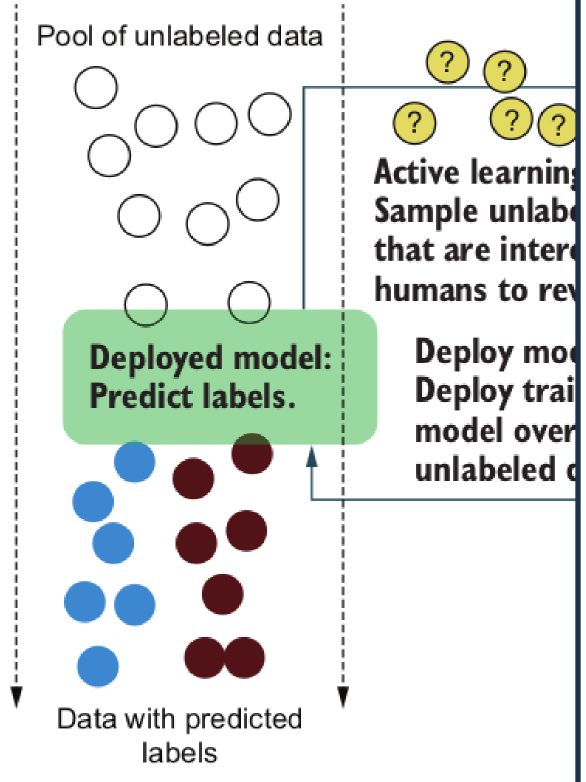


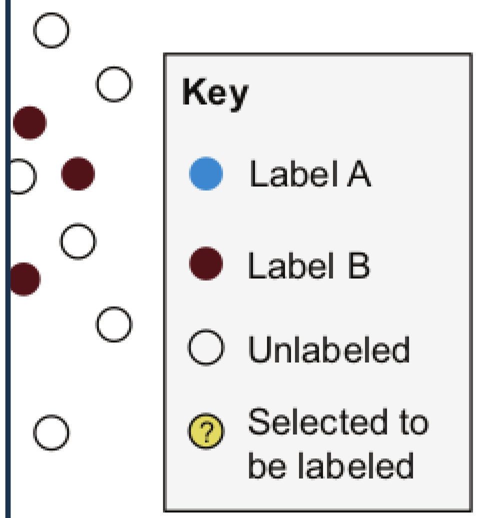
Figure 1.1 A mental model of the hu

Uncertainty sampling

Sampling unlabeled data points near decision boundary of your model. Sampling highest uncertainty first.

Sample No.	uncertainty
1	80%
2	40%
3	30%
4	50%
5	15%

order by uncertainty



**source: Human in the loop machine learning, Robert M.*

Introduction to active learning sampling

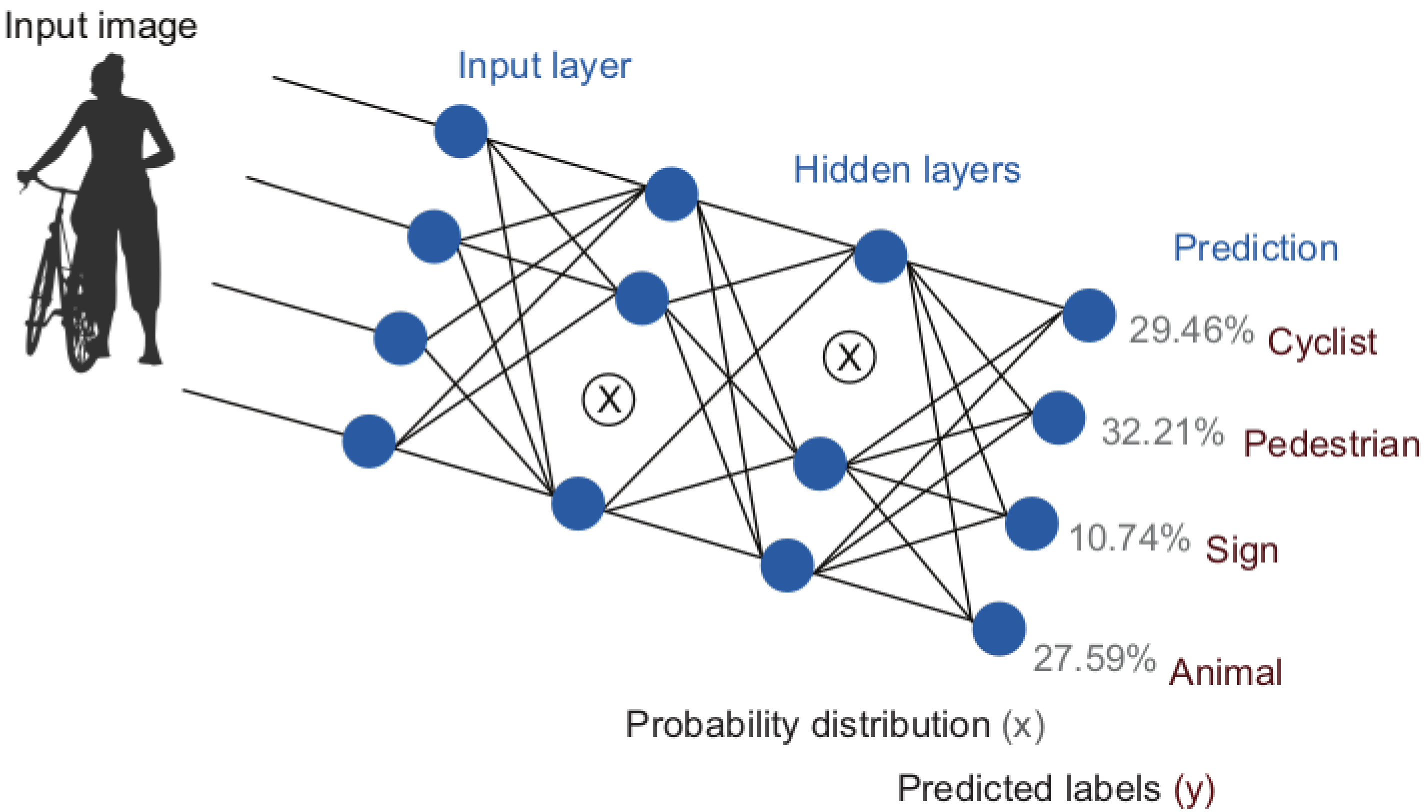
Uncertainty sampling

Multiple techniques possible for uncertainty sampling.

Uncertainty sampling by using dropout layers

- Multiple inference iterations N per sample equivalent of ensemble of models

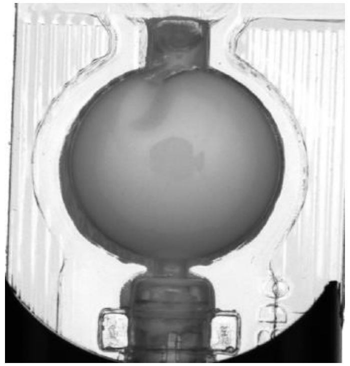
- Variation in predictions determines the uncertainty score

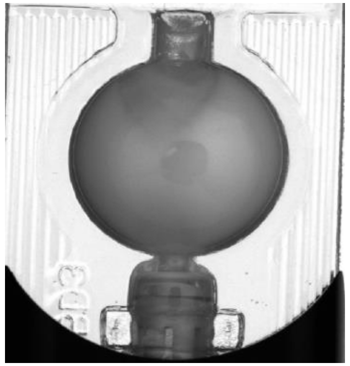


****source: Human in the loop machine learning, Robert M.***

Experiment design and results

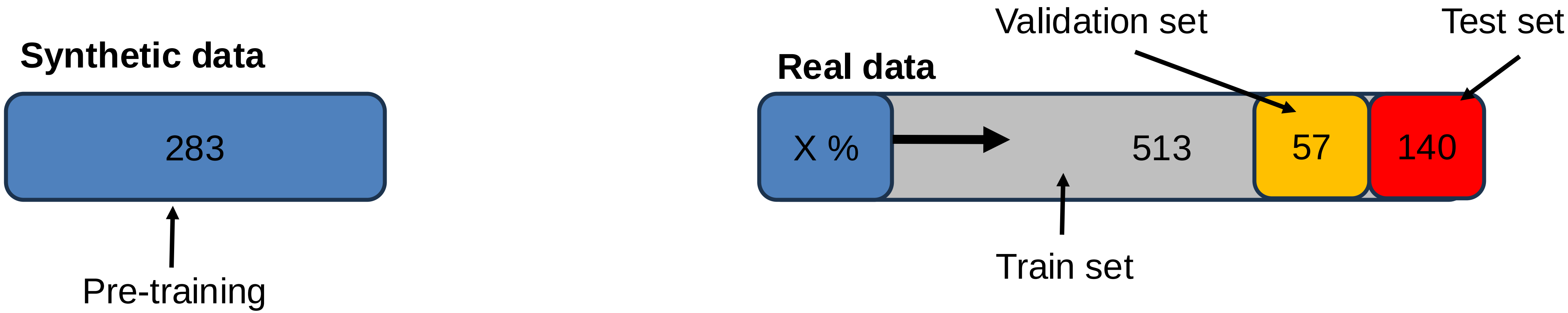
Performance using synthetic hot-start and active sampling





- Synthetic dataset only contain front/back views **283 images**
- Filter real dataset on front/back views **713 images**

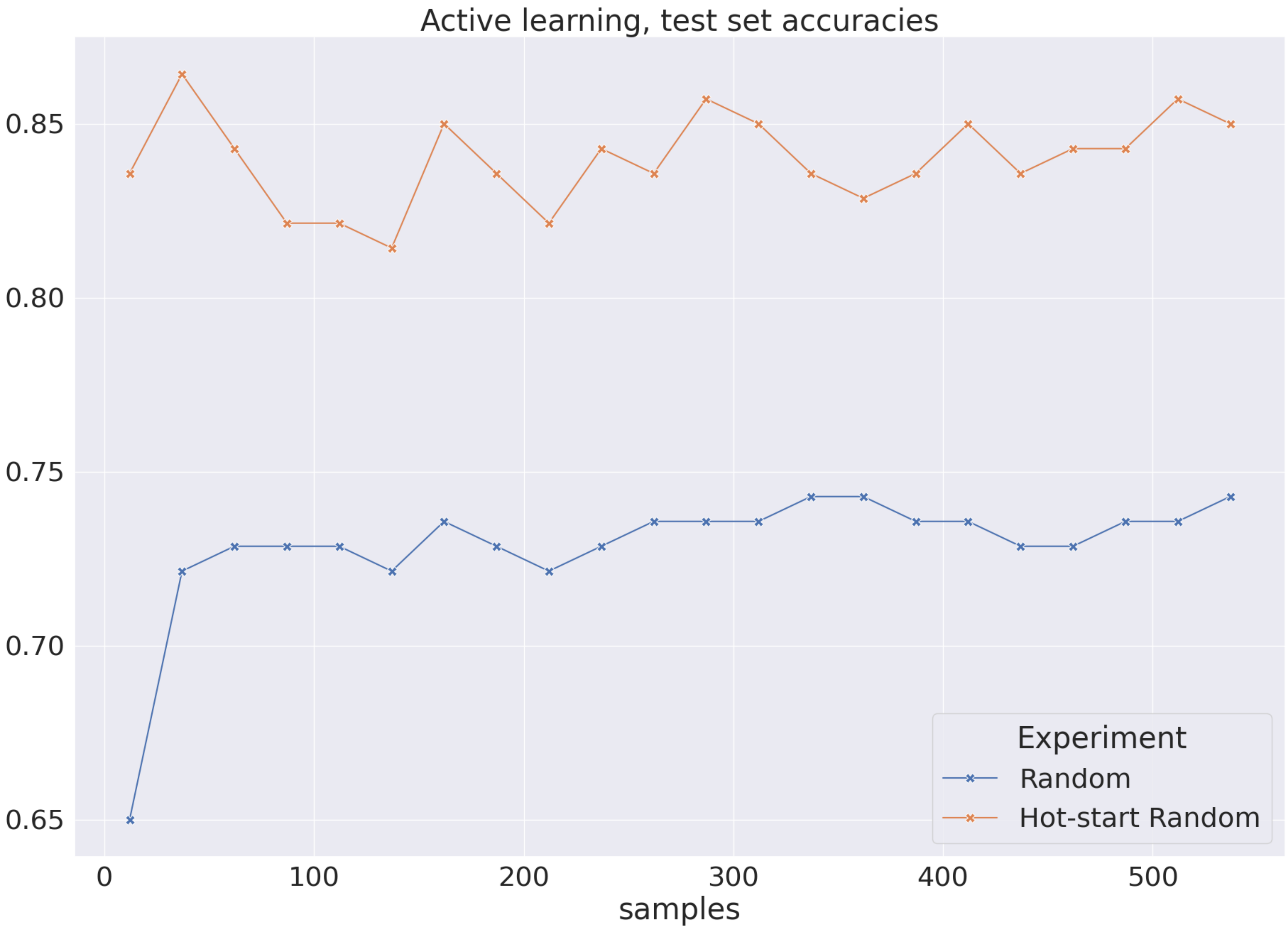
1. 20 samples per **<u>defect type</u>** of real data for test set, **140 images**
2. randomly sampled 10% of remaining real data as validation data, **57 images**
3. remaining real data as pool, **513 images**
   - Initial training size 2.5% => **12 images**
   - Each iteration Increase training size 5% => **25 images**
4. Sampling for step 3 random or using uncertainty method



Experiment design and results

Clear difference between synthetic hot-start and training from scratch

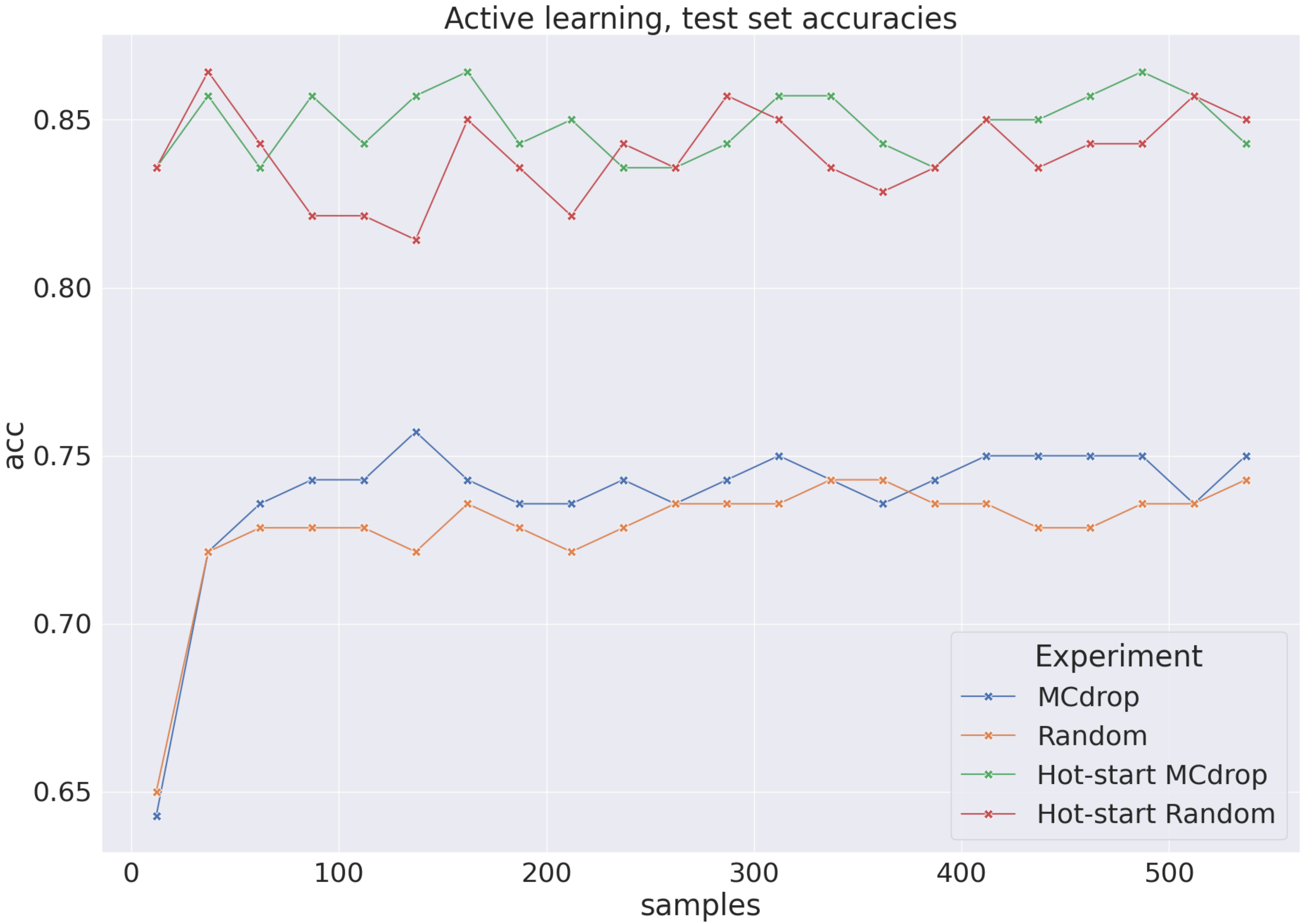
#train samples	Difference %	Difference #samples
12	18.5	26
37	14.3	20
513	11	15



Experiment design and results

Uncertainty sampling doesn't provide relevant increase in accuracy.

For some train set size random sampling is even better



Experiment design and results

Qualitative analysis influence
Synthetic hot-start at
iteration=0, 12 samples

#train samples	Difference %	Difference #samples
12	**18.5**	**26**
37	14.3	20
513	11	15

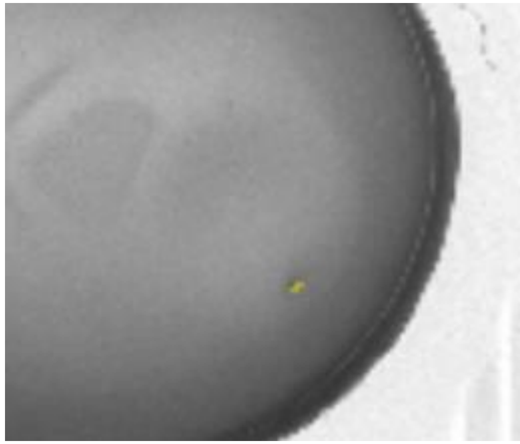
3 images where annotation and prediction don't overlap but very close (iou=0)

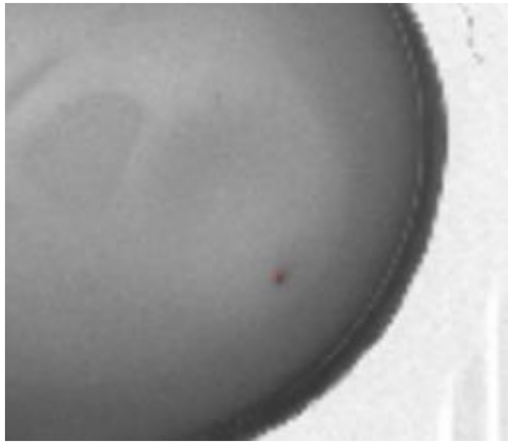
Other wrong predictions mostly tiny dot defects or small line-like defects

- **Predictions wrong/correct here?**
- **Annotations not perfect**

Wrong predicted images
WITH synthetic hot-start

Annotation | **Prediction**





Experiment design and results

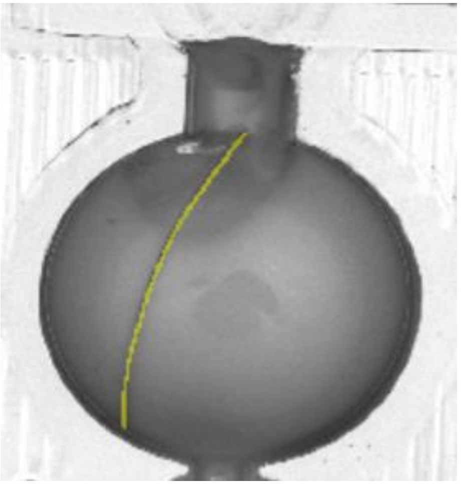
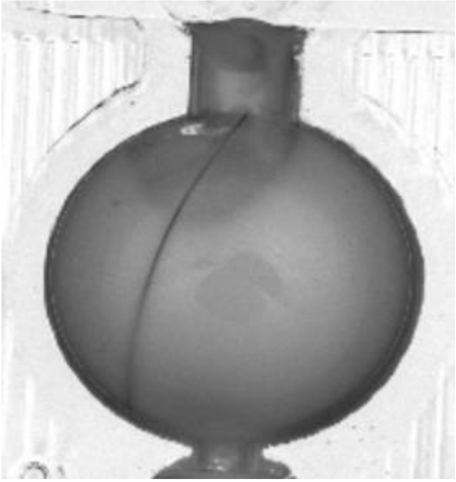
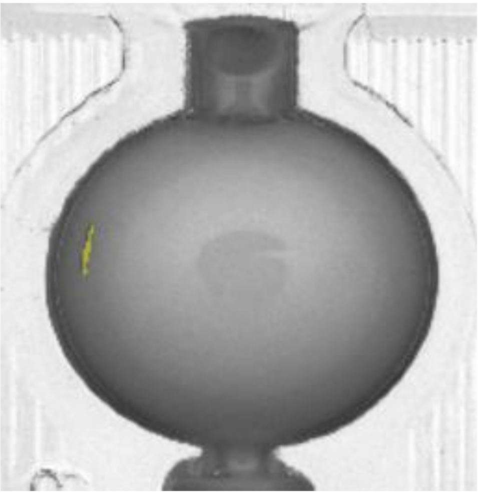
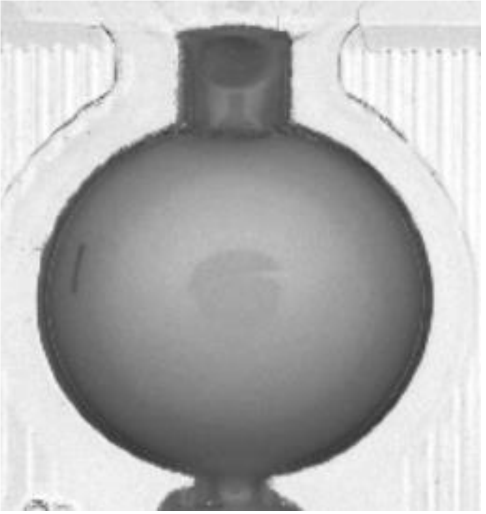
Qualitative analysis influence
Synthetic hot-start at
iteration=0, 12 samples

#train samples	Difference %	Difference #samples
12	**18.5**	**26**
37	14.3	20
513	11	15

12 images with
with long, medium sized line-like defects

line-like defects are not present in synthetic data, larger line-like defects were correctly predicted.

Wrong predicted images
<u>WITHOUT</u> synthetic hot-start

Annotation	**Prediction**
	
	

Experiment design and results

Wrong predicted images
WITHOUT synthetic hot-start

Qualitative analysis influence
Synthetic hot-start at
iteration=0, 12 samples

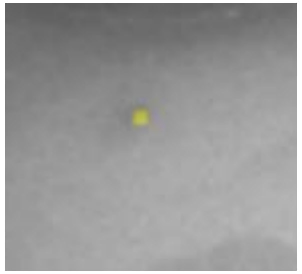
#train samples	Difference %	Difference #samples
12	**18.5**	**26**
37	14.3	20
513	11	15

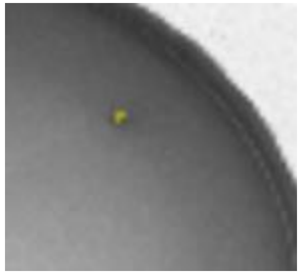
13 images with
small dot-like defects

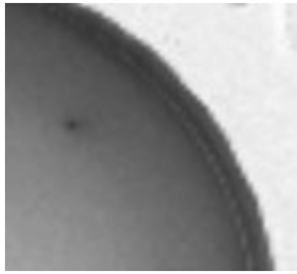
Are missed completely

Annotation

Prediction







Experiment design and results

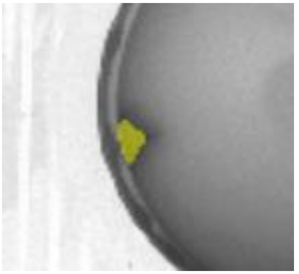
Qualitative analysis influence
Synthetic hot-start at
iteration=0, 12 samples

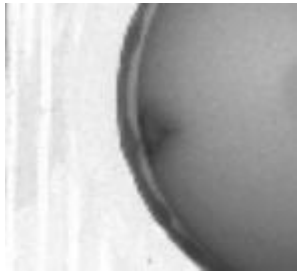
#train samples	Difference %	Difference #samples
12	**18.5**	**26**
37	14.3	20
513	11	15

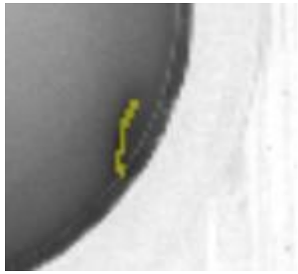
7 images with
larger irregular structured defects

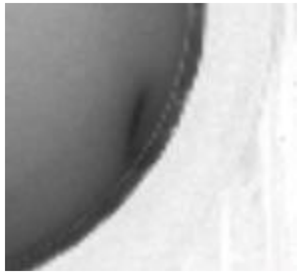
Wrong predicted images
<u>WITHOUT</u> synthetic hot-start

Annotation | **Prediction**









Conclusions

Gap between synthetic hot-start and without is **11-18.5%**
Some defect shapes and grey-pixel values are quite uncommon. Test set is made of 20 samples of each **defect type** = kind of material included inside bulb. This does not equal the shape and pixel values seen by camera.

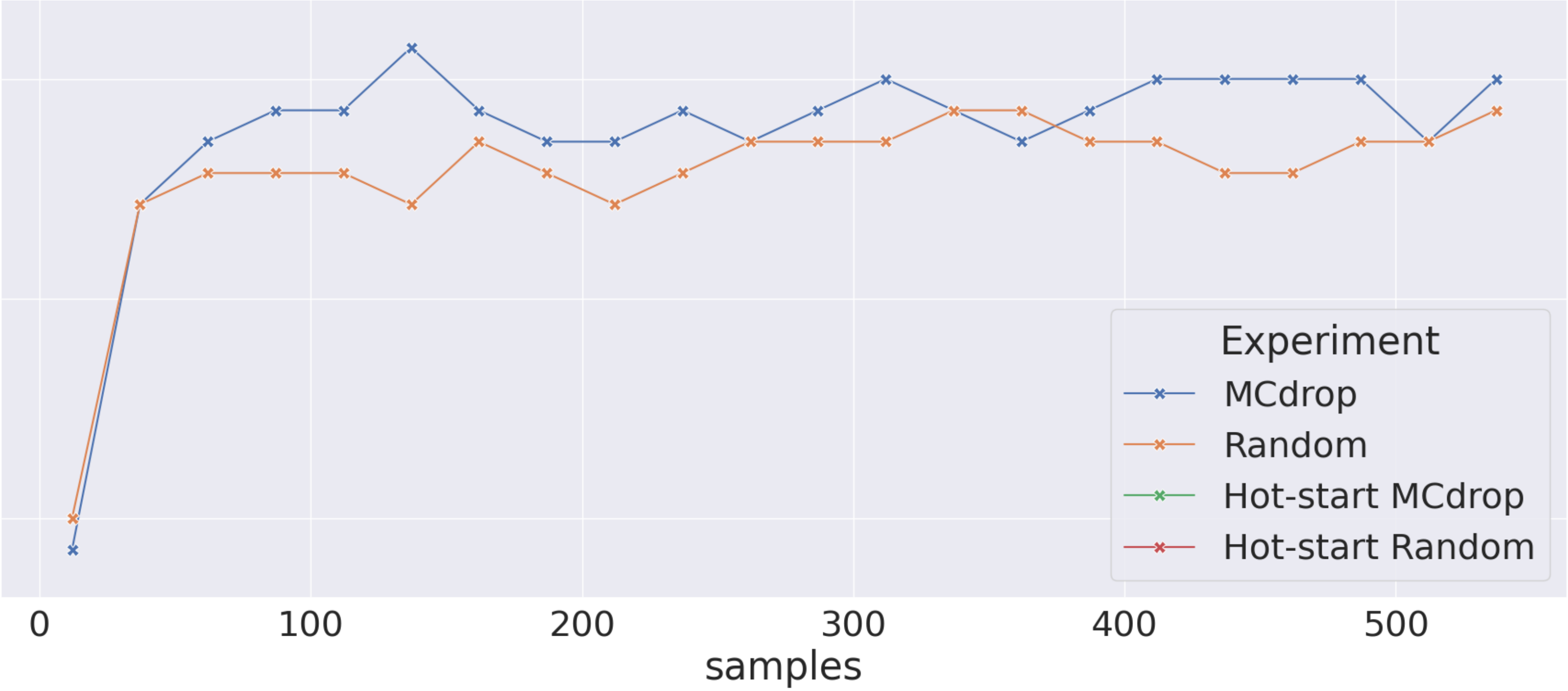
Synthetic data introduces extra variability that allows predictions to be more accurate in general.

#train samples	Difference %	Difference #samples
12	**18.5**	**26**
37	14.3	20
513	11	15

Conclusions

Uncertainty sampling doesn't produce statistically relevant better results, from looking at the shape of defects qualitatively this could be because of two reasons

- The real samples of train set don't include that much variety in shape/types. Active sampling in general will have little effect.
- Uncertainty estimation with a dropout layer is not a good estimation. Other active sampling methods might perform better



Conclusions

Collecting a real dataset of **513 samples** with defects included from different materials is a luxury. Production line needs to be stopped.

Using only synthetic data and **12 samples of real data** respectable accuracies can be achieved on a representative test set. The only information used from the real test data to generate synthetic defects was the shape (dot-like or line-like).
Random perturbations are generated.

This makes our method of synthetic hot-start a valid approach in achieving most of the variability in defect shapes/pixel values as a starting point without affecting production.