

INSTITUTE OF COMPUTER
SCIENCES
Master in Artificial Intelligence and Data
Science

Universitätsstr. 1 D–40225 Düsseldorf



Heinrich Heine
Universität
Düsseldorf

AI-based fluorescent labeling for cell line development

Hanna Pankova

Master thesis

Date of issue: 01. April 2022
Date of submission: 29. August 2022
Reviewers: Prof. Dr. Markus Kollmann
Dr. Wolfgang Halter

Erklärung

Hiermit versichere ich, dass ich diese Master thesis selbstständig verfasst habe. Ich habe dazu keine anderen als die angegebenen Quellen und Hilfsmittel verwendet.

Düsseldorf, den 29. August 2022

Hanna Pankova

Abstract

Cell line development is an expensive and time-consuming process, however that is the most modern approach for producing the proteins needed in pharmaceuticals.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Notation	1
2	Background	2
2.1	Biology	2
2.2	Imaging	2
2.3	ML	2
3	Prediction of cell organelles	3
3.1	Overfitting	3
4	Crops combination	4
5	Predicting cell organelles	5
5.1	Loss	5
5.2	Nuclei	5
5.3	Golgi Apparatus	5
6	Downstream metrics	7
6.1	Nuclei segmentation	7
6.2	ER-segmentation	10
6.3	Golgi	10
7	Stability and generalizability	11
7.1	Stability	11
8	Drift Detection	12
8.1	Embeddings visualization	12
8.2	Drift detection vs Outliers detection	12
8.3	Drift Detection	12
8.4	Practical application of MMD	12
9	Summary	13

List of Figures	14
------------------------	-----------

List of Tables	14
-----------------------	-----------

1 Introduction

1.1 Motivation

1.2 Notation

2 Background

2.1 Biology

2.2 Imaging

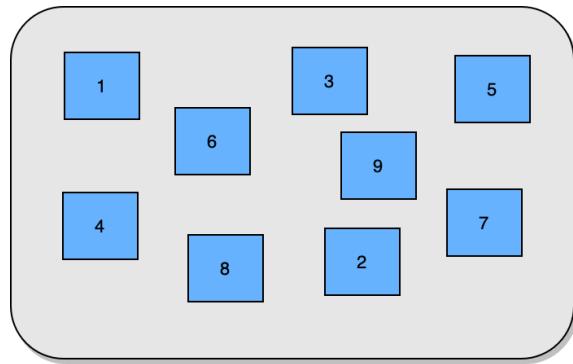


Figure 1: Way in which photos of the well-plate were taken

2.3 ML

3 Prediction of cell organelles

3.1 Overfitting

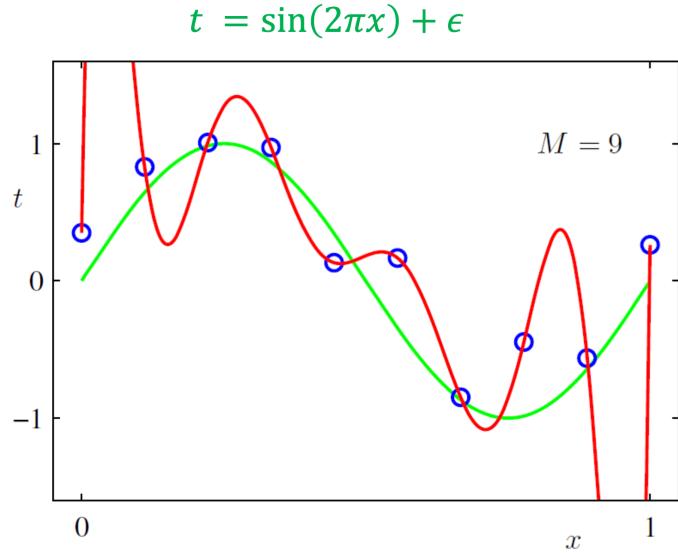


Figure 2: Overfitting

(Bishop book)

3.1.1 Early-stopping

3.1.2 Regularization

3.1.3 Network reduction

3.1.4 Expansion of the training data

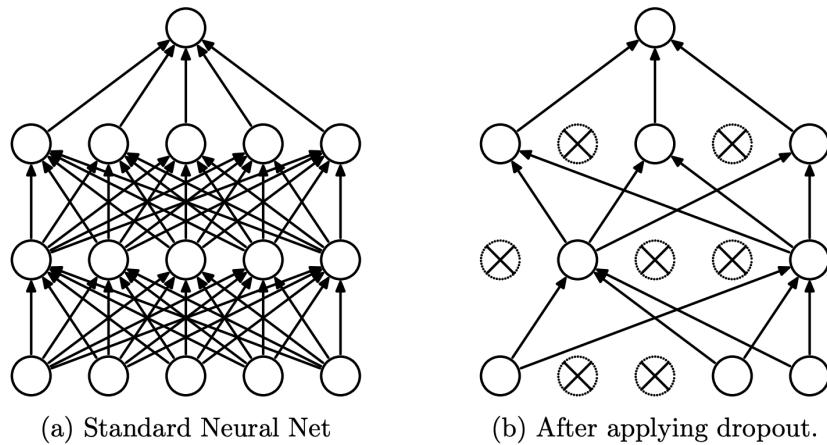


Figure 3: Dropout

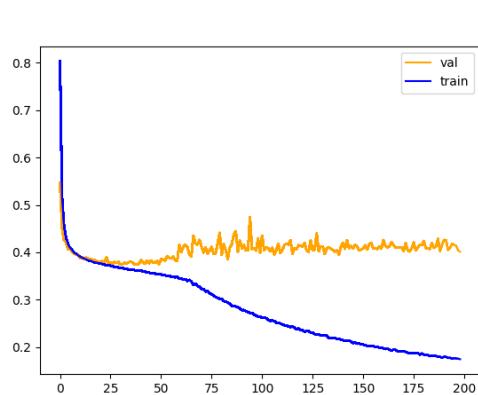


Figure 4: Not regularized

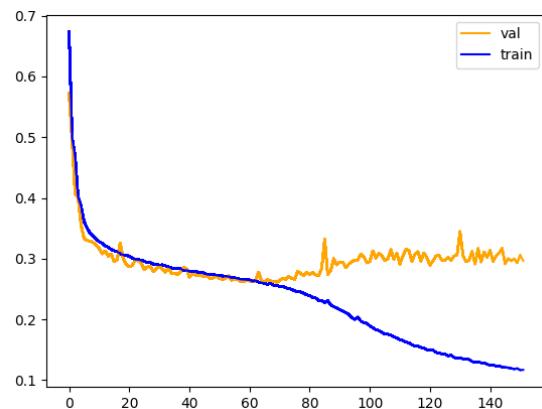


Figure 5: Regularized

4 Crops combination

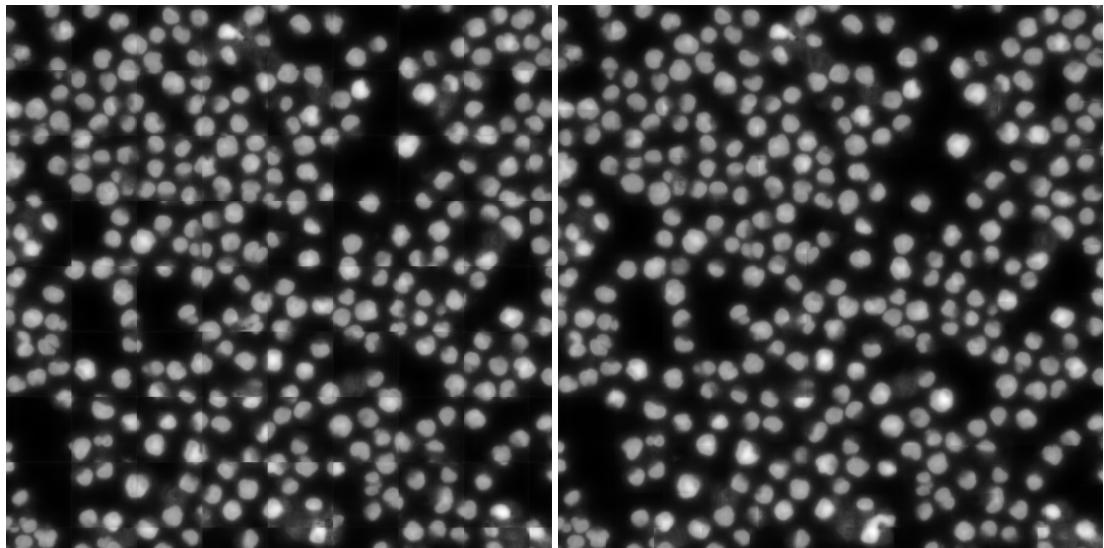


Figure 6: No overlap

Figure 7: 30 pixels overlap

5 Predicting cell organelles

5.1 Loss

5.2 Nuclei

5.3 Golgi Apparatus

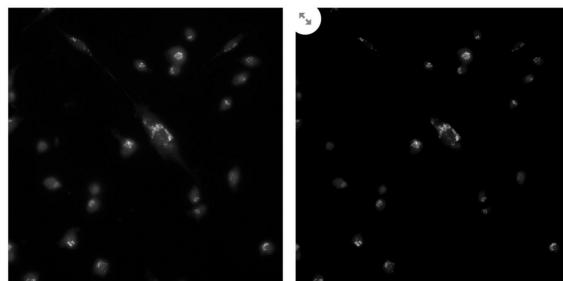


Figure 8: Golgi enhancement

5.3.1 Rolling ball algorithm

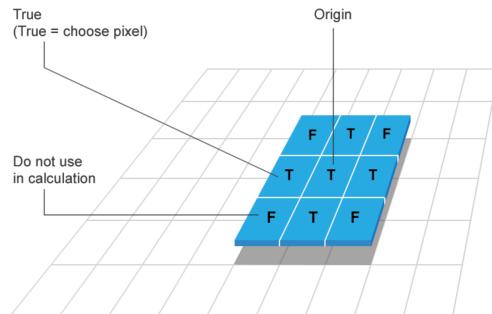


Figure 9: Structuring Element

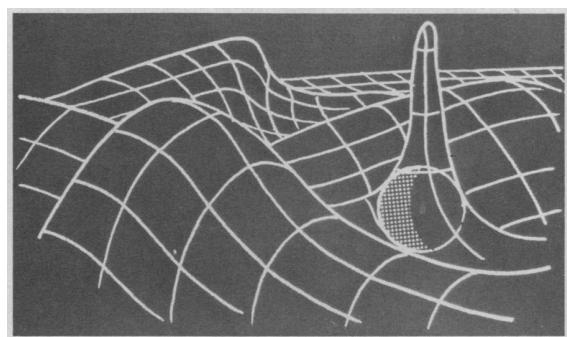


Figure 10: Rolling Ball

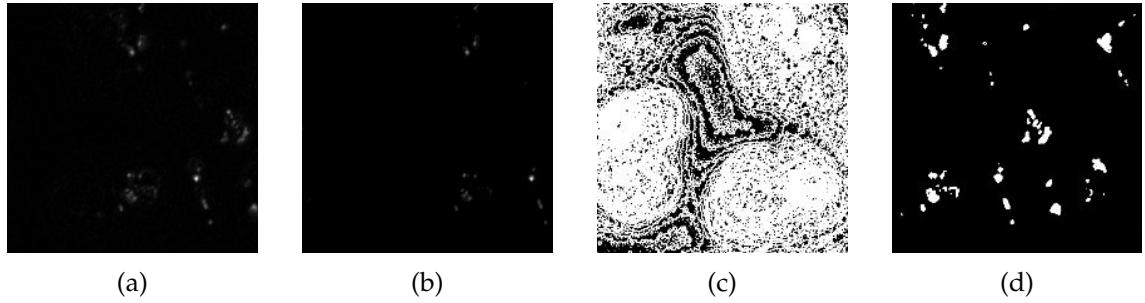


Figure 11: (a) Vanilla pre-processing with automatic background removal algorithm only; (b) Additional clipping of lower intensities after vanilla pre-processing; (c) masked or subfigure (a); (d) mask of subfigure (b)

6 Downstream metrics

6.1 Nuclei segmentation

6.1.1 Challenge

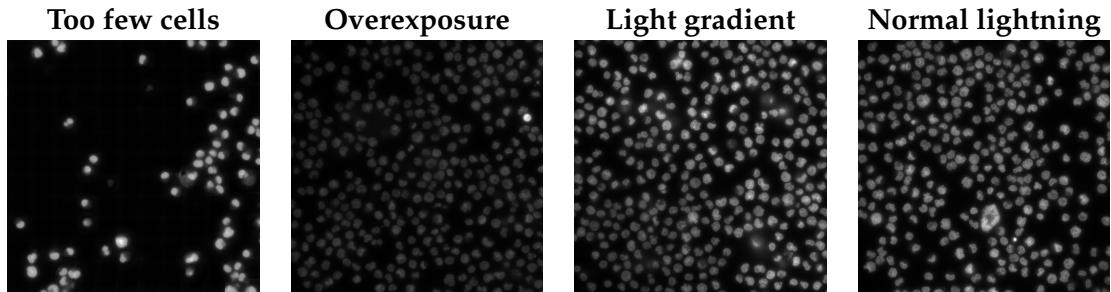


Figure 12: Different lightning conditions

Local Threshold	Global Threshold
0.3 sec	17 sec

Table 1: Threshold timing

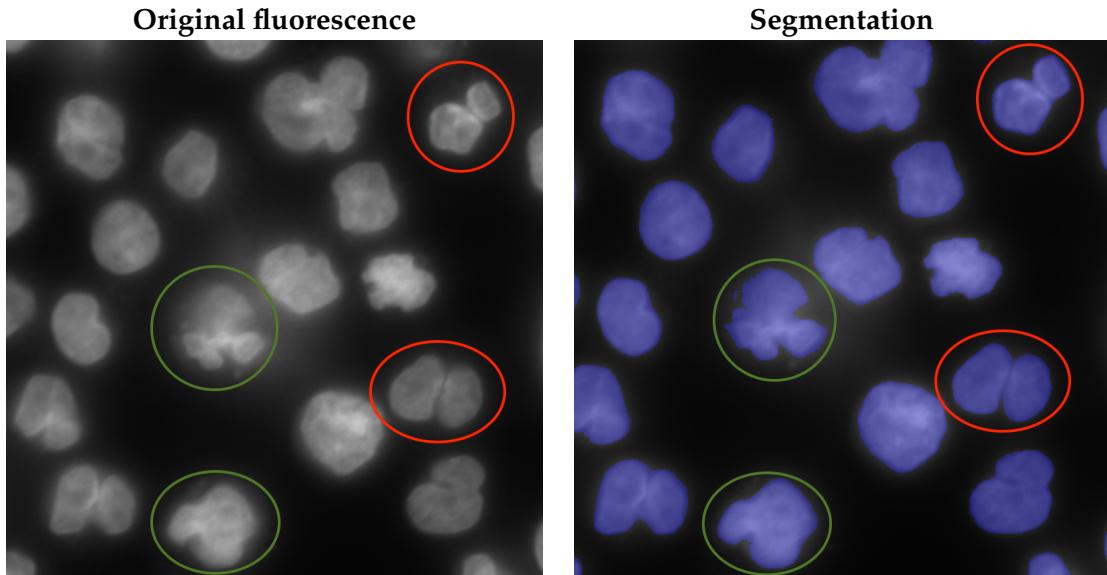


Figure 13: Closely located cells

6.1.2 Thresholding

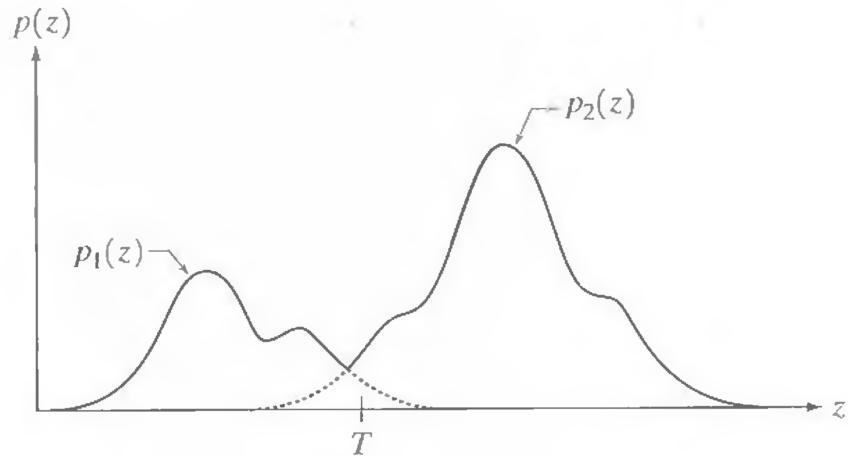


Figure 14: Histogram as a probability density function

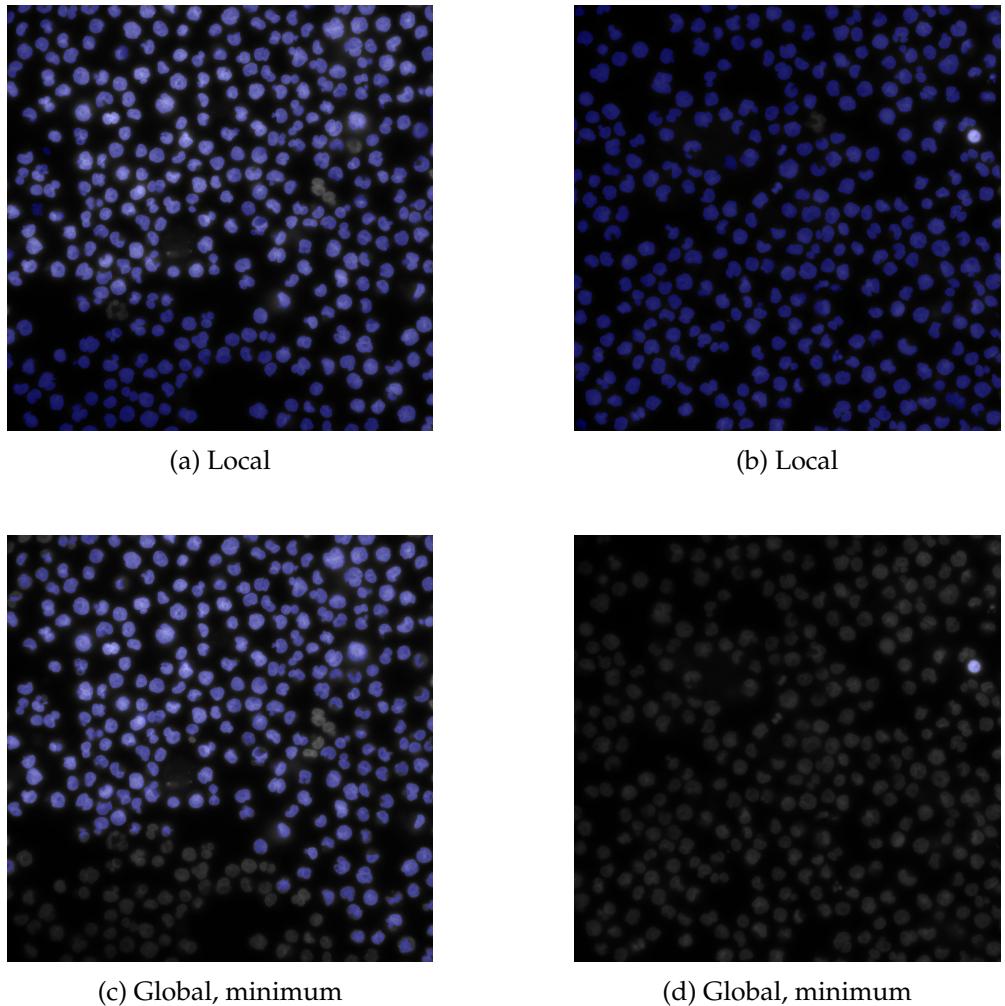


Figure 15: Local vs. Global thresholding

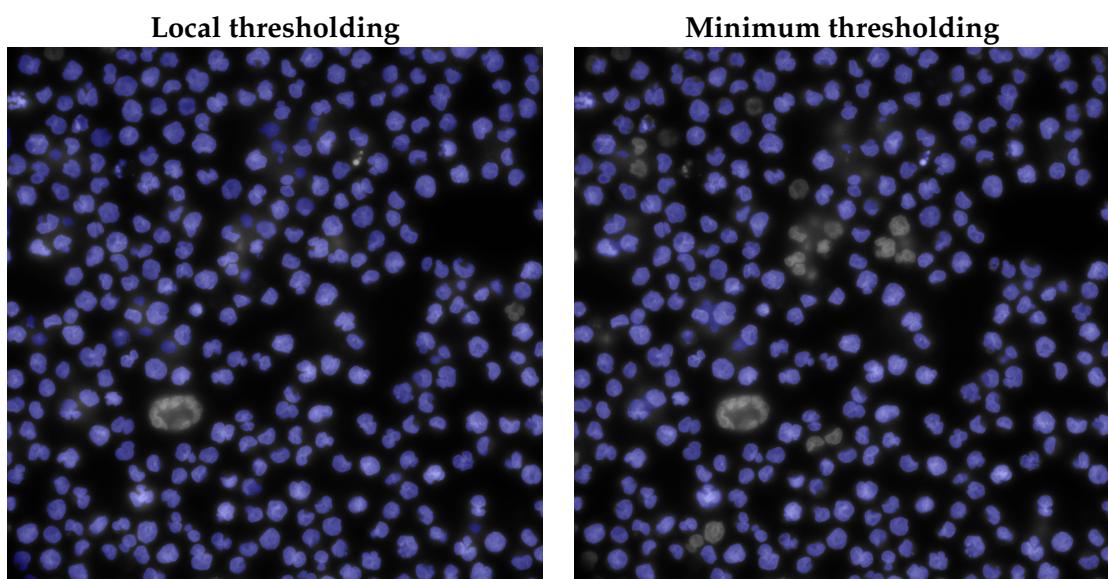


Figure 16: Local vs. Global thresholding (normal conditions)

6.1.3 Overall algorithm

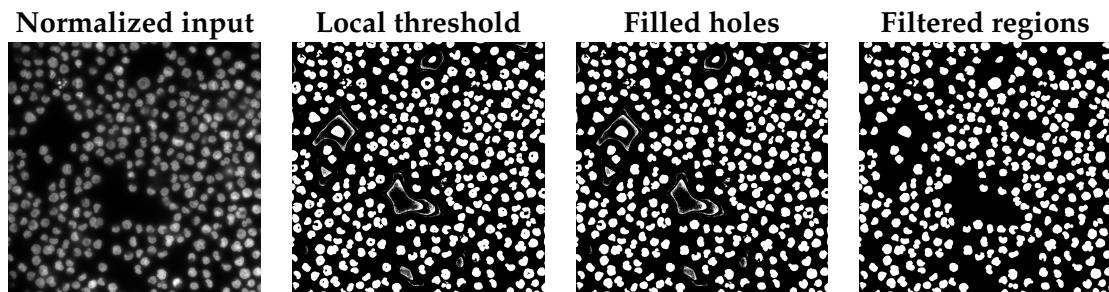


Figure 17: Fluorescence segmentation

6.2 ER-segmentation

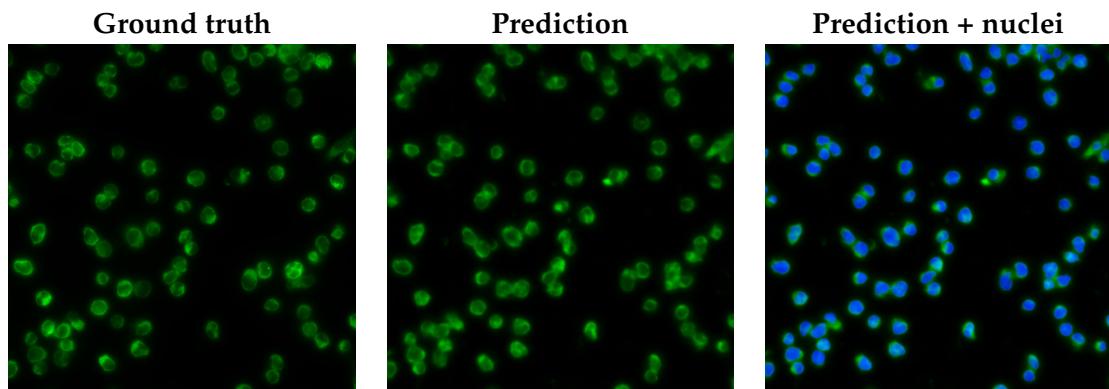


Figure 18: ER prediction

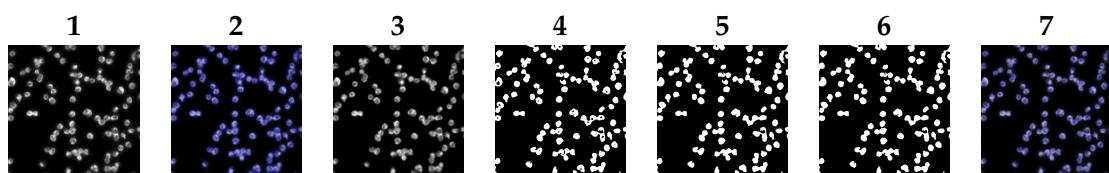


Figure 19: ER prediction

6.3 Golgi

7 Stability and generalizability

7.1 Stability

7.1.1 Gaussian Noise

7.1.2 Defocus Blur

7.1.3 Gaussian Blur

7.1.4 Brightness

7.1.5 Contrast

8 Drift Detection

8.1 Embeddings visualization

8.2 Drift detection vs Outliers detection

8.3 Drift Detection

8.3.1 Kernel methods and two-sample testing

8.3.2 Maximum Mean Discrepancy

8.4 Practical application of MMD

8.4.1 Online MMD

9 Summary

List of Figures

1	Way in which photos of the well-plate were taken	2
2	Overfitting	3
3	Dropout	4
4	Not regularized	4
5	Regularized	4
6	No overlap	5
7	30 pixels overlap	5
8	Golgi enhancement	5
9	Structuring Element	6
10	Rolling Ball	6
11	(a) Vanilla pre-processing with automatic background removal algorithm only; (b) Additional clipping of lower intensities after vanilla pre-processing; (c) masked or subfigure (a); (d) mask of subfigure (b)	7
12	Different lightning conditions	7
13	Closely located cells	8
14	Histogram as a probability density function	8
15	Local vs. Global thresholding	9
16	Local vs. Global thresholding (normal conditions)	9
17	Fluorescence segmentation	10
18	ER prediction	10
19	ER prediction	10

List of Tables

1	Threshold timing	8
---	----------------------------	---