IDENTIFICATION OF PATTERNS IN MICROSCOPY IMAGES OF BIOLOGICAL SAMPLES USING EVOLUTION STRATEGIES

Daniela M. Borgmann [(a)], Julian Weghuber [(b)], Susanne Schaller [(c)], Jaroslaw Jacak [(d)], Stephan M. Winkler [(e)]

[(a, c, e)] University of Applied Sciences Upper Austria
Bioinformatics Research Group
Softwarepark 11, 4232 Hagenberg, Austria

[(b)] University of Applied Sciences Upper Austria
Department of Bio- and Environmental Technology
Stelzhamerstraße 23, 4600 Wels, Austria

[(d)] Johannes Kepler University Linz
Institute of Applied Physics
Altenbergerstraße 69, 4040 Linz, Austria

[(a)] daniela.borgmann@students.fh-hagenberg.at, [(b)] julian.weghuber@fh-wels.at ,
[(c)] susanne.schaller@fh-hagenberg.at, [(d)] jaroslaw.jacak@jku.at, [(e)] stephan.winkler@fh-hagenberg.at

ABSTRACT

In this paper we describe an algorithm based on evolutionary algorithms for determining patterns in images of biological samples (especially living cells) generated using the micro-patterning assay approach. In order to identify these patterns it is necessary to identify symmetric grids in nanoscale microscopy images.

The algorithm presented in this paper is based on evolution strategies (ES): After downsampling the image using a correlation based approach for estimating the optimal downsampling rate, initial grids are constructed which are repeatedly evaluated and mutated for creating new candidates from which the best ones are promoted to the next generation. In the experimental section of this paper we analyse the performance of several ES strategies for identifying optimal grids in several images of biological samples.

Keywords: bioinformatics, evolution strategy, µ-patterning assay

1. INTRODUCTION: USING THE µ-PATTERNING ASSAY FOR THE DETECTION OF PROTEIN-PROTEIN-INTERACTIONS IN LIVING CELLS

The cellular membrane of biological cells is an important integral part of cellular interaction processes. The membrane itself represents a physical barrier between intracellular and extracellular spaces and hereby generates a tiny reaction volume – the biological cell. In order to guarantee the proper function of a cell, interactions between the extracellular environment and the intracellular space are necessary.

In many cases interactions take place with the help of proteins (receptors), which can be found integrated or on the surface of cell membranes. The communication between these receptors is regulated by so-called messenger-substances; for example the binding of insulin to the insulin receptor activates this receptor, which leads to decreased blood sugar levels. Another example can be found in the context of cell-growth: Binding of epidermal growth factor (EGF) to the EGF-receptor (EGFR) can stimulate cell proliferation.

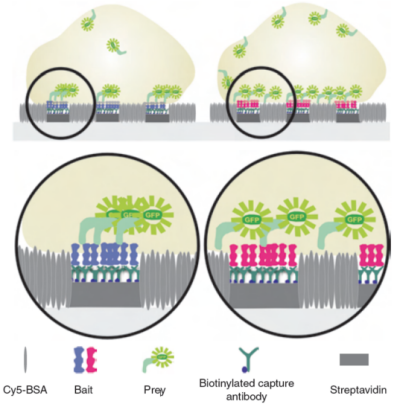

Figure 1: Basic principle of µ-patterning assay

A faulty regulation or an incorrect transmission of EGF can thus lead to cancer.

Protein-protein interactions, especially in cell membranes, represent a key element for many cellular processes. Nowadays, the number of available methods for the detection and quantification of protein-protein interactions in living cells is limited. Some of them entail serious drawbacks like insufficient sensitivity or a high number of false-positive or false-negative results.

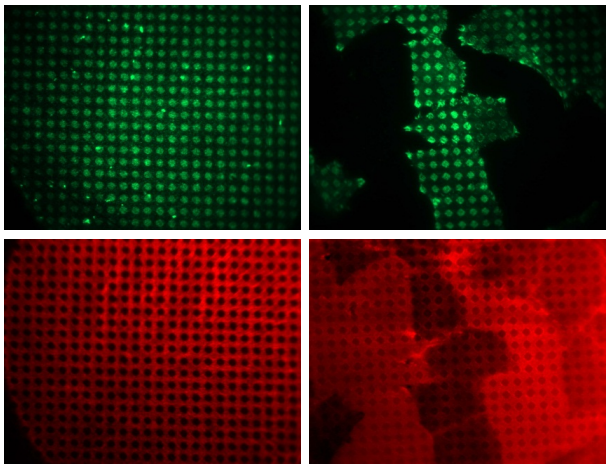

Figure 2: Left column: fluorescently labelled bait antibody control, right column: bait-prey redistribution in living cells on µ-biochips; top: green channel images (interaction detection channel), bottom: red channel images (BSA-Cy5 grid)

A very sensitive method to detect protein-protein interactions in living cells is called micro-patterning assay ((Schwarzenbacher et al., 2008) and (Weghuber et al., 2010a) and (Weghuber et al., 2010b)). Figure 1 illustrates the basic principle of this technique: In order to identify and quantify interactions between a fluorophore-labelled protein (prey) and a membrane protein (bait) *in vivo*, a specific ligand to the exoplasmic domain of the bait is arranged in micropatterns on a glass surface. The intermediate gaps are passivated with BSA. When cells expressing the bait are plated on such surfaces, the bait follows the antibody patterns. To address bait-prey interactions, the lateral distribution of fluorescently tagged prey is analysed and compared with the antibody/BSA micropatterns. Interaction leads to pronounced co-patterning, whereas no interaction yields homogeneous prey-distribution. To detect the interactions a TIRF-based fluorescent microscopy imaging system is best suited, since it allows a strong background reduction from peripheral cell volume.

In order to benefit from the advantages of the µ-patterning technique, appropriate software is needed. Currently this is not the case. During the analyses of bait-prey protein interactions a large amount of data is collected, which has to be analysed. This analysing process cannot be carried out with the possibilities provided in an acceptable time. Thus, new algorithmic solutions are necessary.

2. RESEARCH GOAL: OPTIMIZATION OF GRIDS FOR DETECTING PATTERNS IN BIOLOGICAL IMAGES

The goal of the research work presented in this paper is to develop an algorithm that is able to automatically identify grid structures in images of biological samples labelled by µ-patterning.

The main idea is the analysis of two different kinds of images which represent in combination one biological sample:

- Green light intensity images represent fluorescently labelled protein spots, known as patterns, which are to be analysed.
- Red light intensity images represent the lattices which were used to produce the micro patterned glass plate.

Real world examples for input data analysed in this research work can be seen in Figure 2.

Within these images the patterns have to be separated from areas that represent the lattices. Thus, what is needed is a method that is able to identify symmetric grids within the red channel images and uses these grids to identify pattern areas in the corresponding green channel images.

3. DATA PREPROCESSING

3.1. Correlation Based Optimal Downsampling

In order to decrease the runtime consumption of further image analysis steps (including the identification of grid structures), the analyzed red channel images are downsampled (Lin and Dong 2006) using a correlation threshold θ:

Starting with downsampling rate (*dsr*) 2, the *dsr* is constantly increased until the correlation (Rodgers and Nicewander 1988) of the downsampled image and the original image becomes less than θ.

3.2. Transformation of Greyscale to Binary Images

In order to distinguish pattern areas from lattice, pixels below a pre-defined greyscale intensity level are transformed to white representing grid pixels, others to black ones. Additionally, pixels with intensity above a pre-defined threshold are also transformed to black. This leads to binary images as the one exemplarily shown in Figure 3.

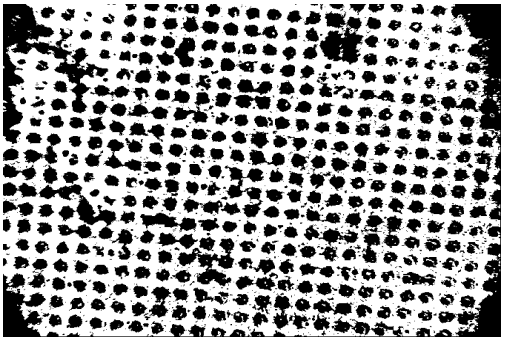

Figure 3: Binary representation of a red channel image

4. IDENTIFICATION OF GRID STRUCTURES IN IMAGES OF BIOLOGICAL SAMPLES BASED ON EVOLUTION STRATEGIES

In this section we describe an approach for identifying grid structures in images using evolution strategies: First, an initial grid is identified, which is repeatedly evaluated and mutated for creating new candidates from which the best ones are promoted to the next generation.

4.1. Evolution Strategies

Evolution strategies (ESs), beside GAs the second major representative of evolutionary computation, were developed since the 1960s, primarily by a German research community around Rechenberg and Schwefel at the Technical University of Berlin, and have been extensively studied in Europe (see for example (Rechenberg 1973) and (Schwefel 1994)).

As it is an evolutionary algorithm, the optimization process based on ES is executed by applying operators in a loop, i.e., main operations are applied on the solution candidates repeatedly until a given termination criterion is met. Similar to GAs, an ES works with a population of individuals; each individual is characterized by its parameter vector which is used to calculate the individual's fitness value. In every step of the algorithm's execution (that is, in each generation), the old population is replaced by a new one. Still, there are differences between GAs and ESs, especially in the form of their genotypes, the calculation of the fitness values and the operators (mutation, recombination and selection):

- ESs most frequently use real-coding of design parameters, they model the organic evolution at the level of individual's phenotypes; the representation used is a fixed-length real-valued vector, each position in the vector corresponds to a feature of the individual.
- Whereas GAs use mutation only for avoiding stagnation, mutation is the main reproduction operator in evolution strategies: Each component of the parameter vector is mutated individually in each generation. Small mutations are more likely than big ones, the standard mutation distribution being the Gaussian mutation ($N(0,\sigma)$).
- In addition to mutation, recombination can be used to create a new individual (a "child") out of two "parents", too. Recombining two ES solution candidates means calculating the geometric average of the parents' parameter vectors.
- In contrast to nature and GAs, the selection of ESs works in a totally deterministic way: In each generation only the best individuals survive.

Examples for mutation and recombination in the context of ESs are shown in Figure 4:

Figure 4: Exemplary solution candidates and the effect of genetic operations in ES

In each generation of an ES algorithm, λ children are produced by μ parent individuals; by selection, the best children are chosen and become the parents of the next generation. Typically, parent selection in ES is performed uniformly randomly, with no regard to fitness; survival in ESs simply saves the μ best individuals, which is only based on the relative ordering of fitness values.

Basically, there are two selection strategies for ESs:

- The (μ,λ)-strategy: μ parents produce λ children; the best μ children are selected and form the next generation's parents.
- The (μ+λ)-strategy: If this selection model, also called the "plus-selection", is applied, μ parents produce λ offspring; parents and children together form a pool of potential new parents, and the best μ individuals are selected from this pool to become the next generation's parents.

The main procedure steps of the execution of ES is summarized and graphically shown in Figure 5.

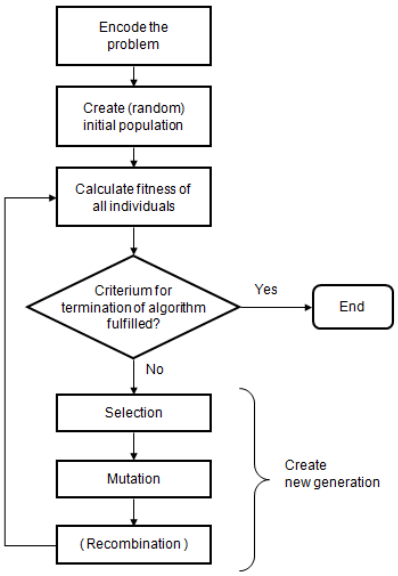

Figure 5: Workflow of the standard evolution strategy (ES) algorithm

Rechenberg proposed a heuristic for the adaptation of the mutation variance, the so-called 1/5 success rule (which was originally proposed for the special case of a

(1+1) ES; similar rules have also been stated for other ES-variants): The quotient of the number of the successful mutants (those that improve the population's quality) to all mutants should be about 1/5. If this quotient is greater than 1/5, then the mutation variance should be increased; if the quotient is less than 1/5 (which means that less than 20% of the mutations produce better mutants), the mutation variance should be reduced.

Additionally, this ES workflow is extended as proposed by Schwefel (Schwefel 1994) so that there is an individual mutation strategy parameter for each parameter of the solution candidate. Thus, the mutation strength of each feature is also optimized during the evolutionary process.

4.2. Solution Candidates Representing Grids

A solution candidate is represented as a composition of four parameters and its quality measure. As shown in Figure 6, the four parameters are:

- The grid's *deflection*, the inclination of the grid referring to the image orientation,
- the grid's *width*, i.e., the distance between two gridlines,
- the *horizontal offset* of a reference vertical grid line,
- the *vertical offset* of a reference horizontal grid line.

4.3. Evaluation of Grid Solution Candidates

In order to obtain a robust and reproducible comparison between various grid solution candidates an evaluation function has been defined; this evaluation function calculates the quality of a solution candidate comparing the binary image, retrieved from the input grid image, with the solution candidate itself.

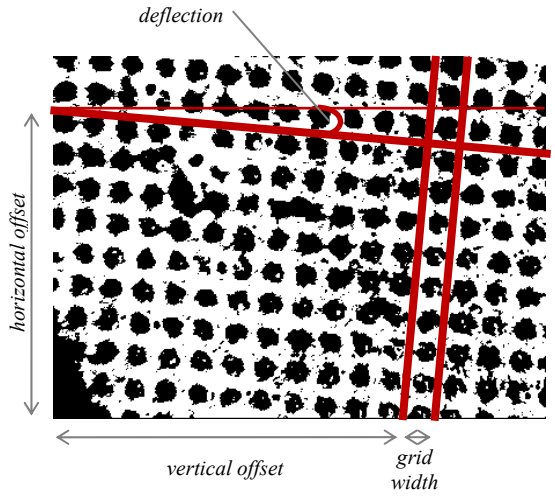

Figure 6: Parameters of a grid candidate

4.3.1. Computation of the Whole "Expected" Grid Defined by a Grid Solution Candidate

As a solution candidate is only represented by four parameters, it is clearly necessary to expand the solution candidate of the size of the initial image and thus retrieve the whole grid defined by this parameter combination.

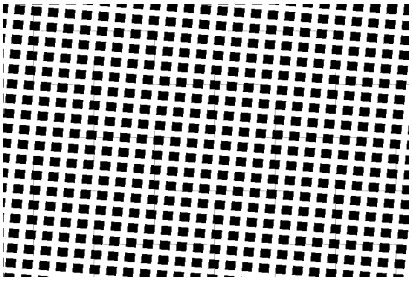

Figure 7: Exemplary expected grid

This clamping process is done by calculating all gridlines appearing in the initial image and furthermore (*gridwidth * gridwidthTolerance)* additional lines on the left and the right side of the original line. These additional lines are calculated to be aware of inaccuracies in the original grid picture and therefore to avoid faulty calculations of intensity and contrast values in the analyses steps later on.
The result of a fully clamped expected grid can be seen in Figure 7. The result is a binary image with white-coloured grid lines and black-coloured patterns.

4.3.2. Comparison with Initial Binary Image

The actual quality of a solution candidate (*grid*) is computed by the comparison of each pixel of the expected grid image (*expected(grid)*) with the corresponding pixel of the binary representation of the original image (*binary*). The number of positive comparisons is summed up over all pixels *p* and divided by the total number of pixels *N*. The retrieved value is within the range [0, 1].

$$N(binary) = |\{p \mid p \in binary\}| \quad (1)$$

$$P(grid, binary) = |\{expected(grid)[p] = binary[p]\}| \quad (2)$$

$$quality(grid, binary) = \frac{|P(grid, binary)|}{|N(binary)|} \quad (3)$$

4.4. Identification of Initial Solution Candidates

In order to start the identification of grid structures we determine initial solution candidates which are calculated from the binary image. The first step is to define three gridlines (two vertical and one horizontal) in order to be able to complete the grid definition. A gridline itself is composed of an offset-value, which determines the position within the image, and its

deflection. A gridline can be evaluated by calculating the percentage of white pixel-values compared to the binary image (as defined in Equation 5).

- First, starting from some area A within the image (e.g., the right upper quarter of the image), all possible vertical gridlines in this are evaluated:

$$\forall\,(x \in x_A, deflection \in [-0.2; +0.2]):\\ l = verticalGridline(x, deflection);\\ quality(l) = eval(l, binaryImage) \tag{4}$$

$$eval(gridline, binaryImage) = \frac{|\{p | p \in gridline \wedge binaryImage[p] = white\}|}{|\{p \in gridline\}|} \tag{5}$$

  The best gridline with offset x and deflection def ($l_{opt}(x_{opt}, def_{opt})$) so found in A is the starting point for the construction a complete grid definition.

- Second, the best next parallel gridline $l_{opt}2$ to l_{opt} is computed by iteratively decreasing (or increasing) x (starting at x_{opt}) and evaluating the so created lines with the same deflection def_{opt} until the next local optimum in the set of possible vertical gridlines is found.

- Finally, all horizontal gridlines in A with deflection $-def_{opt}$ are calculated:

$$\forall\,(y \in y_A):\\ l = horizontalGridline(y, -def_{opt});\\ quality(l) = eval(l, binaryImage) \tag{6}$$

  The best of these is selected as $l_{opt}3$.

These three lines (l_{opt}, $l_{opt}2$, and $l_{opt}3$) define an initial grid candidate identified in the area A.

This initial grid construction heuristic is performed five times in different starting areas of the given image; the best so found solution candidate is subsequently optimized by the ES algorithm.

5. EXPERIMENTAL RESULTS

In order to demonstrate ability of the here presented approach to identify optimal grids in biological microscopy images, 8 exemplary samples have been selected. These samples represent microscopy images of living cells and their size is 1344 by 1024 pixels.

In the following tables we summarize the qualities achieved for these samples for downsampling tolerances 0.8 and 0.9 as well as varying ES parameter settings (plus selection as well as mutation strategy adaptation as described previously were applied in all tests). In Figure 8 we show the average improvements achieved by one specific ES configuration for the given samples (w.r.t. the initial qualities); examples of grid structures found for one of the chosen microscopy images are shown in Figure 9.

	Quality		
Image Sample	**initial**	**optimized**	**Improvement [%]**
1	0.749	0.788	5.213%
2	0.726	0.822	13.335%
3	0.701	0.778	11.110%
4	0.605	0.804	32.738%
5	0.651	0.767	17.779%
6	0.584	0.733	25.424%
7	0.696	0.699	0.459%
8	0.548	0.570	3.978%
Mean Improvement [%]			**13.754%**

Table 1: Results achieved with downsampling tolerance 0.8, μ = 5, λ = 10, and 100 ES iterations

	Quality		
Image Sample	**initial**	**optimized**	**Improvement [%]**
1	0.682	0.770	12.964%
2	0.709	0.803	13.208%
3	0.649	0.776	19.599%
4	0.572	0.808	41.124%
5	0.543	0.719	32.484%
6	0.525	0.619	17.813%
7	0.561	0.661	17.781%
8	0.594	0.643	8.263%
Mean Improvement [%]			**20.404%**

Table 2: Results achieved with downsampling tolerance 0.9, μ = 5, λ = 10, and 100 ES iterations

	Quality		
Image Sample	**initial**	**optimized**	**Improvement [%]**
1	0.749	0.786	4.968%
2	0.726	0.816	12.424%
3	0.701	0.774	10.523%
4	0.605	0.801	32.232%
5	0.651	0.763	17.202%
6	0.584	0.729	24.867%
7	0.696	0.698	0.418%
8	0.548	0.562	2.593%
Mean Improvement [%]			**13.154%**

Table 3: Results achieved with downsampling tolerance 0.8, μ = 5, λ = 10, and 40 ES iterations

	Quality		
Image Sample	**initial**	**optimized**	**Improvement [%]**
1	0.682	0.769	12.825%
2	0.709	0.800	12.784%
3	0.649	0.775	19.446%
4	0.572	0.804	40.435%
5	0.543	0.717	32.145%
6	0.525	0.619	17.818%
7	0.561	0.661	17.697%
8	0.594	0.642	8.065%
Mean Improvement [%]			**20.152%**

Table 4: Results achieved with downsampling tolerance 0.9, μ = 5, λ = 10, and 40 ES iterations

Image Sample	Quality		Improvement [%]
	initial	optimized	
1	0.749	0.788	5.174%
2	0.726	0.822	13.278%
3	0.701	0.778	11.042%
4	0.605	0.802	32.543%
5	0.651	0.763	17.231%
6	0.584	0.732	25.402%
7	0.696	0.698	0.406%
8	0.548	0.565	3.031%
Mean Improvement [%]			**13.513%**

Table 5: Results achieved with downsampling tolerance 0.8, $\mu = 3$, $\lambda = 15$, and 40 ES iterations

Image Sample	Quality		Improvement
	initial	optimized	
1	0.682	0.769	12.820%
2	0.709	0.784	10.491%
3	0.649	0.776	19.607%
4	0.572	0.807	41.078%
5	0.543	0.758	39.594%
6	0.525	0.684	30.274%
7	0.561	0.661	17.731%
8	0.594	0.641	8.018%
Mean Improvement			**22.452%**

Table 6: Results achieved with downsampling tolerance 0.9, $\mu = 3$, $\lambda = 15$, and 40 ES iterations

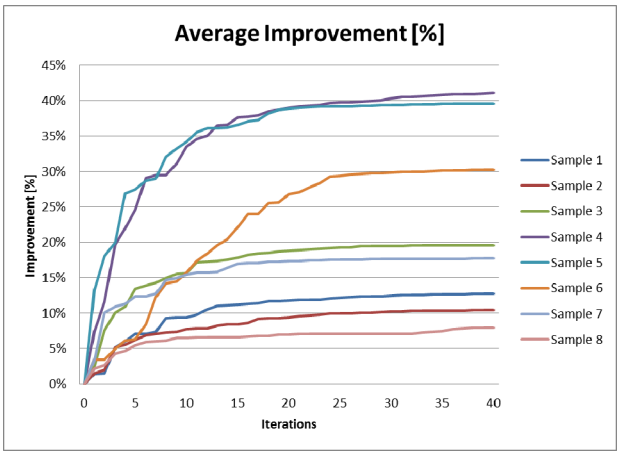

Figure 8: Average improvement during the evolutionary process for the 8 given samples (compared to the initial solution candidate) achieved with $\mu = 3$, $\lambda = 15$, and downsampling tolerance 0.9.

6. CONCLUSION

In this paper we have described an evolutionary process with an initial construction heuristic that is able to automatically identify grids in microscopy images of biological samples. In future work this approach shall be included in a fully automated bioinformatical image analysis framework that is designed for biomedical research.

ACKNOWLEDGMENTS

The work described in this paper was done within the research project µProt/DETECTOR sponsored by Basic Research Program of the University of Applied Sciences Upper Austria.

REFERENCES

Lin, W., Dong, L., 2006. Adaptive Downsampling to Improve Image Compression at Low Bit Rates. *IEEE Trans. Image Processing* 15:2513-2521.

Rechenberg, I., 1973. *Evolutionsstrategie*. Friedrich Frommann Verlag.

Rodgers, J. L., Nicewander, W. A. 1988. Thirteen ways to look at the correlation coefficient. *The American Statistician* 42:59-66.

Schwarzenbacher, M., Kaltenbrunner, M., Brameshuber, M., Hesch, C., Paster, W., Weghuber, J., Heise, B., Sonnleitner, A., Stockinger, H., Schütz, G. J., 2008. Micropatterning for quantitative analyses of protein-protein interactions in living cells. *Nature methods 5:*1053-1060.

Schwefel, H.-P., 1994. *Numerische Optimierung von Computer-Modellen mittels der Evolutions-strategie*. Basel: Birkhäuser Verlag.

Weghuber, J., Sunzenauer, S., Plochberger, B., Brameshuber, M., Haselgrübler, T., Schütz, F. J., 2010. Temporal resolution of protein-protein interactions in the live-cell plasma membrane. *Analytical and Bioanalytical Electrochemistry 397:* 3339-3347.

Weghuber, J., Brameshuber, M., Sunzenauer, S., Lehner, M., Paar, C., Haselgrübler, T., Schwarzenbacher, M., Kaltenbrunner, M., Hesch, C., Paster, W., Heise, B., Sonnleitner, A., Stockinger, H., Schütz, G.J., 2010. Detection of protein-protein interactions in the live cell plasma membrane by quantifying prey redistribution upon bait micro patterning. *Methods in Enzymology, 472:* 133-151.

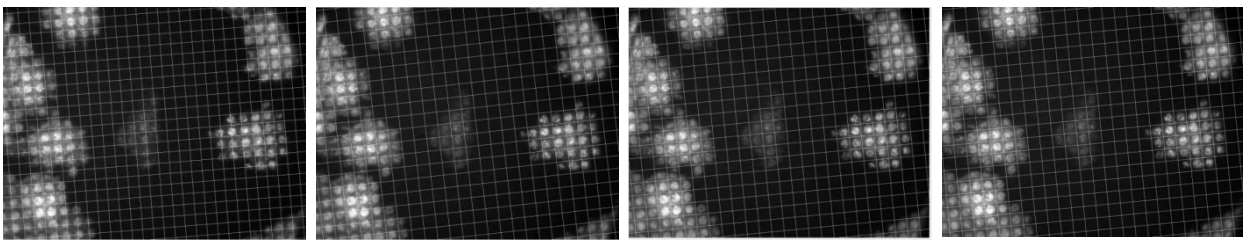

Figure 9: Grid improvement during the ES process ($\mu = 5$, $\lambda = 10$, downsampling tolerance 0.9; first image: initial solution (quality 0.55), second image after 10 iterations (quality 0.62), third after 20 iterations (quality 0.71), fourth after 40 iterations (final grid quality: 0.72)