A Multi-Objective Optimization Algorithms

Algorithms for multi-objective optimization aim to simultaneously optimize two or more objectives. In contrast to single-objective optimization, multi-objective optimization provide a set of points known as Pareto optimal set. In a Pareto optimal set, different solutions represent the trade-off solutions between conflicting objectives. A variety of mathematical techniques have been developed to obtain Pareto optimal solutions [MA04]. However, such techniques present several limitations, e.g. the susceptible to the continuity and shape of the Pareto front. Furthermore, they usually only generate one element of the Pareto optimal set per execution. These limitations gave rise to more flexible and easy-to-use metaheuristics. Currently, among the most popular metaheuristics are evolutionary multi-objective optimization algorithms [Co20]. They fall into three main categories: pareto-based, indicator-based, and decomposition-based.

Pareto-based algorithms use a selection mechanism based on Pareto ranking. The core idea of Pareto ranking is to rank the population according to Pareto optimality and include a density estimator to maintain as many different solutions in the population as possible. Since its proposal by Goldberg [Go89], several algorithm implementations based their selection scheme on Pareto Ranking (e.g. Multi-Objective Genetic Algorithm (MOGA) [FF93] or Nondominated Sorting Algorithm (NSGA) [SP94]). However, those first Pareto-based algorithms had a high computational complexity as well as a lack of elitism. Elitism is the concept of retaining the best solutions obtained in the population to prevent the evolutionary operators from destroying well-performing solutions. NSGA-II [De02], the predecessor of NSGA, introduced elitism, a more efficient ranking schema, and a crowded comparison estimator as density estimator. Despite the well-known limitations of the crowd comparison operator and the rapid increase of non-dominated solutions, when dealing with more than three objectives, NSGA-II is still used by researchers today [Is17].

Indicator-based algorithms were introduced to tackle the increasing number of non-dominated solutions in multi-objective problems with more than three objectives. The modified selection scheme of indicator-based algorithms incorporates a performance measure(e.g. hypervolume or ϵ indicator). Zitzler2004 [ZK04] provided an algorithmic framework for incorporation of performance indicators into the selection scheme in Indicator Based Evolutionary Algorithm (IBEA). Interest was sparked by the introduction of the S Metric Selection Evolutionary Multiobjective Algorithm (SMS-EMA) [EBN05] that combines the crossover, mutation, and non-dominated sorting of NSGA-II with the hypervolume. However, as SMS-EMA and its modified versions rely on the calculation of the exact hypervolume contributions, it becomes computationally very expensive with the increasing number of objectives [Be09].

The key idea of decomposition-based approaches is transforming a multi-objective problem into a variety of single-objective problems that are solved to generate non-dominated solutions to the original problem. Decomposition methods allow the generation of non-convex and disconnected Pareto front to overcome the limitation of linear aggregation

functions. MOEA/D [ZL07] decomposes a multi-objective problem into several scalar subproblems by scalarizing the objective functions with different weight vectors. Each subproblem is optimized simultaneously using only information from its neighboring subproblems, which leads to a lower complexity than NSGA-II at each generation and easier diversity maintenance. However, MOEA/D has other limitations: the population size is the size of the weight vectors, which leads for many-objective problems to an impractically large number of weight vectors [Is17]. For handling this difficulty, NSGA-III [DJ14] adopts both: decomposition and reference points. Ishibuchi et al. [Is17] showed that the performance of decomposition-based MOEAs (including MOEA/D and NSGA-III) are highly sensitve to Pareto front shapes and can be outperformed on those shapes even by NSGA-II. Furthermore, Ishibuchi et al. [Is17] stated, that for high performance of decomposition-based method: 1) The triangular shape of the pareto front is the same as or similar to the distribution of the weight vectors. 2) The Pareto Front has to be small in comparison with the feasible region in the objective space, and 3) The decision variables should be separable.Nevertheless, decomposition-based approaches are still an active area of research.

Taking into account that our optimization problem only has three objectives, all of the above introduced algorithm are a valid choice. However, we exclude indicator-based methods since our optimization problem does not fall into the area of many-objective problems and therefore we avoid the extra complexity of calculating an indicator. Further, we exclude decompositon-based methods since we cannot fullfill two of the three reasons Ishibuchi et al. [Is17] analyzed as being responsible for high performance. The exclusion of indicator- and decomposition-based approaches leaves us with the pareto-based approaches. NSGA-II compared to MOGA and NSGA implements a more efficient ranking scheme leading to increasing popularity. Despite its age,researchers still apply NSGA-II to numerous problems in the last decade (e.g. [STD11,Dh11,Ma16]), including a counterfactual generating problem with tabular data [Da20]. Furthermore, the performance of NSGA-II has been tested in many competitive studies (e.g. [Is17, Is09, RZR12, GSP16]). Therefore, we decided to use NSGA-II.

B Distance Functions for Images

This section contains the used Image Similarity Measures and the necessary reshaping of those to range $[0, 1]$ and the minimization Problem. Throughout x denotes the original "factualïmage while x' denotes the counterfactual

B.1 Root Mean Squared Error

The Root Mean Squared Error (RMSE) measures the average change per pixel. RMSE values are non-negative and a value of 0 means the image or videos being compared are identical.

$$O_1(x, x') = RMSE(x, x') \tag{4}$$

$$RMSE = \sqrt{\frac{1}{M * N} \sum_{i=0,j=0}^{M-1,N-1} [x(i,j) - x'(i,j)]^2} \tag{5}$$

B.2 Mean Absolute Error

The mean absolute error between two images corresponds to the L_1-norm. It is normalized to the range $[0, 1]$ by dividing through the maximal pixel value of 255.

$$O_1(x, x') = MAE(x, x')/255 \tag{6}$$

$$MAE = \frac{1}{M * N} \sum_{i=0,j=0}^{M-1,N-1} |x(i,j) - x'(i,j)| \tag{7}$$

B.3 Structural Similarity Index

The Structural Similarity Index (SSIM) which in contrast to the plain distance is able to account for temporal and spatial relationships. As the SSIM takes on values between -1 and 1, we scale the values to the rang $[0, 1]$. A value of 1 stands for identical images, while the lower the values the less correlated are the images. By reformulating the scaled SSIM into a minimization problem, we obtain the structural dissimilarity index [SAU19], the distance measure derived from the SSIM.

$$\begin{aligned} O_1(x, x') &= 1 - (\frac{SSIM(x, x') - (-1)}{(1 - (-1))} * (1 - (-1)) + (-1)) \\ &= \frac{1 - SSIM(x, x')}{2} \end{aligned} \tag{8}$$

$$\begin{aligned} SSIM(x, x') &= \frac{1}{M} \sum_{\forall i,j} SSIM(x, x'; i, j) \\ &= \frac{1}{M} \sum_{\forall i,j} l(x, x'; i, j) \\ &\quad * c(x, x'; i, j) \\ &\quad * s(x, x'; i, j) \\ &= \frac{1}{M} \sum_{\forall i,j} (\frac{2\mu_{x(i,j)}\mu_{x'(i,j)} + C_1}{\mu^2_{x(i,j)} + \mu^2_{x'(i,j)} + C_1}) \\ &\quad * (\frac{2\sigma_{x(i,j)}\sigma_{x'(i,j)} + C_2}{\sigma^2_{x(i,j)} + \sigma^2_{x'(i,j)} + C_2}) \\ &\quad * (\frac{\sigma_{x(i,j),x'(i,j)} + C_3}{\sigma_{x(i,j)}\sigma_{x'(i,j)} + C_3}) \end{aligned} \tag{9}$$

The SSIM score of an entire image is computed by moving a sliding window across the image and averaging the local (window-wise) SSIM values across an image. The local SSIM score consists of three parts: the similarity of the local patch luminance $l(x, x')$, the similarity of the local patch contrast $c(x, x')$ and the similarity of the local patch structure $s(x, x')$. The values are calculated by basic statistics. μ_x and $\mu_{x'}$ denote the local sample means of x and x' and σ_x and $\sigma_{x'}$ respectively the standard deviation. $\sigma_{x,x'}$ refers to the cross-correlation of x and x' after removing their means. C_1, C_2 and C_3 are small positive constants to prevent numerical instability by dividing to or by zero. For more information on the SSIM please refer to [ZB09].

B.4 Feature-Based Similarity Index

The Feature-Based Similarity Index(FSIM) [ZGD11] compares the structural and Feature similarity measure between two images.It is based ob phase congruence and gradient magnitude. Thereby $S_L(x)$ denotes the overall similairity and $PC_m(x)$ the phase congruency. Ω denotes the whole spatial image domain. The feature range is between $[0, 1]$, where 1 indicates perfect feature similarity.

$$O_1(x, x') = 1 - FSIM(x, x') \tag{10}$$

$$FSIM = \frac{\sum_{x \in \Omega} S_L(x) * PC_m(x)}{\sum_{x \in \Omega} PC_m(x)} \tag{11}$$

B.5 Information theoretic-based Statistic Similarity Measure

The Information theoretic-based Statistic Similarity Measure (ISSM) [Al19] is a hybrid approach incorporating information theory (Shannon entropy) with a statistic (SSIM) as well as a distinct structural feature provided by edge detection (Canny).

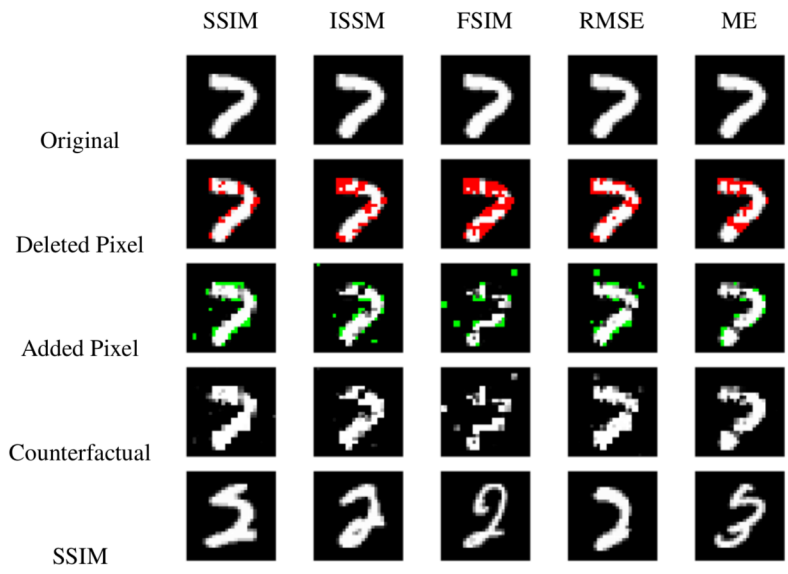
$$O_1(x, x') = 1 - ISSM(x, x') \tag{12}$$

$$ISSM(x, x') = \frac{C(x, x')EHS(x, x')(a + b) + e}{aC(x, x')EHS(x, x') + bEHS(x, x') + cSSIM(x, x') + e} \tag{13}$$

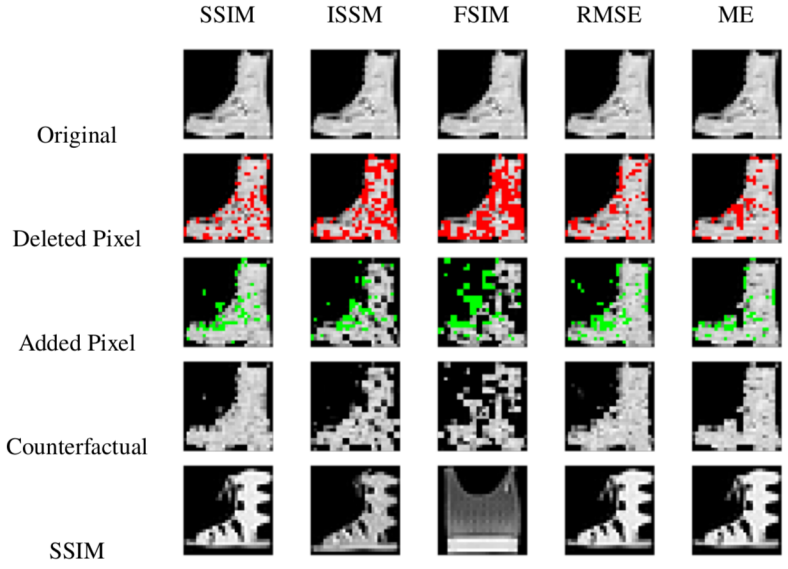
EHS denotes the Entropy-Histogram Similarity, C the canny edge detector and $SSIM$ the structural similarity index. The remaining constants ($a = 0.3$, $b = 0.5$, and $c = 0.7$) are added to balance the quotient and avoid division by zero.

C Additional Results

This section contains additional information, visualizations and results to Section 4. Abb. 5 visualizes the original image and the generated counterfactual for all the tested distances. Abb. 6 show the influence of the mutation operator on an examples. Abb. 7 visualizes the pixel changes obtained by our approach to achieve the conterfactual.

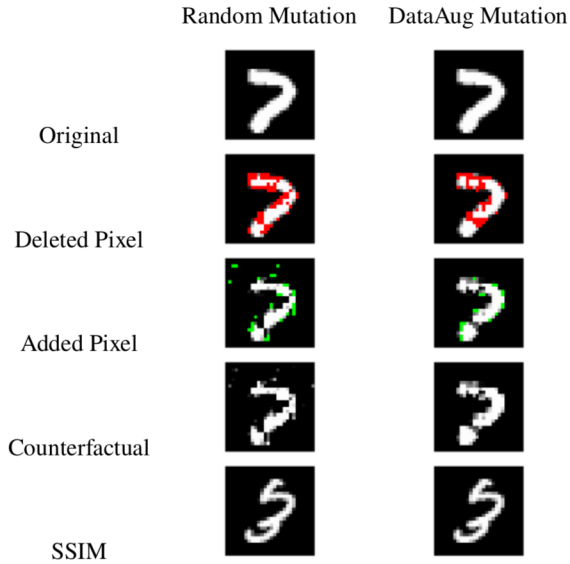


(a) MNIST

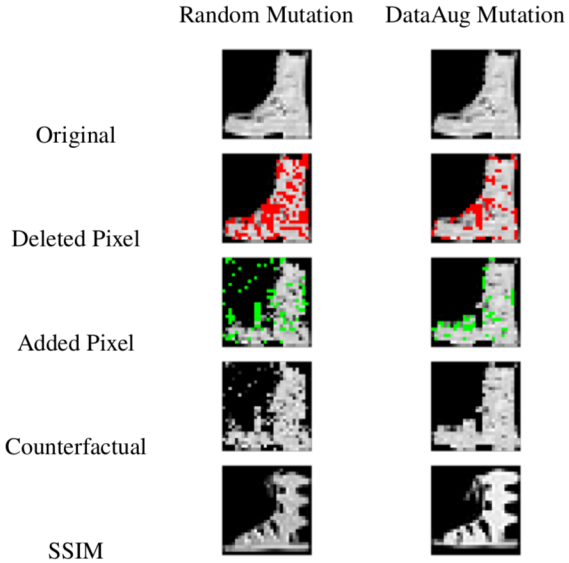


(b) Fashion MNIST

Abb. 5: Visual Comparison of distance measures. Original Image and Counterfactual generated on test image 8030 of MNIST and 9377 of Fashion MNIST. The first row shows the original image, the second row highlights the pixels deleted ($x - x' < 0$) to get to the counterfactual, and the third row shows the added pixels ($x - x' > 0$). The following row visualizes the counterfactual. For reference purposes, the last row shows the image with the lowest SSIM to the counterfactual on the predicted counterfactual class.

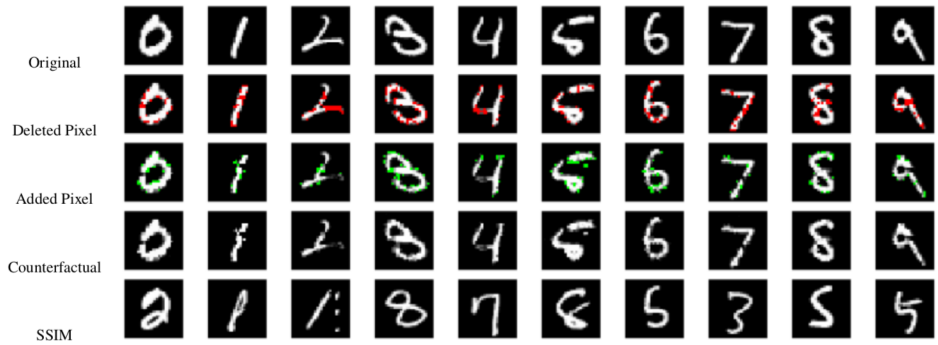


(a) MNIST

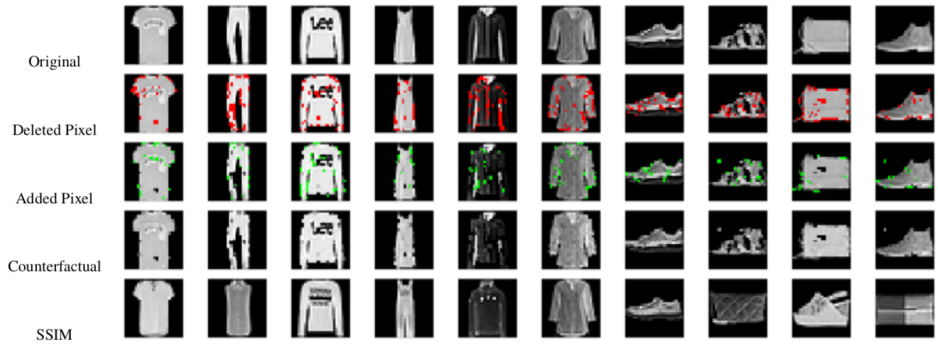


(b) Fashion MNIST

Abb. 6: Visual comparison of mutation operators. Original Image and Counterfactual generated on test image 8030 of MNIST and 9377 of Fashion MNIST. The first row shows the original image, the second row highlights the pixels deleted ($x - x' < 0$) to get to the counterfactual, and the third row shows the added pixels ($x - x' > 0$). The following row visualizes the counterfactual. For reference purposes, the last row shows the image with the lowest SSIM to the counterfactual on the predicated counterfactual class.



(a) MNIST



(b) Fashion

Abb. 7: Visualization of changed pixels for our approach. The first row shows the original image, the second row highlights the pixels deleted ($x - x' < 0$) to get to the counterfactual, and the third row shows the added pixels ($x - x' > 0$). The following row visualizes the counterfactual. For reference purposes, the last row shows the image with the lowest SSIM to the counterfactual on the predicted counterfactual class.