A Genetic Algorithm-Based Solver for Very Large Jigsaw Puzzles Supplementary Material

Dror Sholomon
dror.sholomon@gmail.com

Omid David
mail@omiddavid.com

Nathan S. Netanyahu*
nathan@cs.biu.ac.il

Department of Computer Science, Bar-Ilan University, Ramat-Gan 52900, Israel

1. Introduction

This document provides supplementary material for our paper, "A Genetic Algorithm-Based Solver for Very Large Jigsaw Puzzles". We present here all of the results obtained using our genetic algorithm (GA)-based solver on both public datasets and new large image benchmarks assembled by us. Moreover, we present additional figures, which better illustrate the puzzle solution, in general, and the crossover procedure, in particular. All images in this PDF file are included in a relatively high resolution. Thus, readers of the electronic version of the PDF are encouraged to zoom-in, in order to better appreciate the solutions' accuracy. Following is the description of all figures and tables.

We ran our solver 10 times on every image in every set, except for the 22,834-piece puzzle sets (due to time constraints, we have managed to complete only 2 runs per image for the latter case, before the submission deadline). All puzzle pieces were of 28×28 pixels. For every set we calculated three types of an "average result": "average best", "average worst" and "average average", including an "average standard deviation". For example, the "average best" of an entire set is computed by obtaining the best result, per each image, over 10 runs on that image, and averaging over the number of images in the set. For a set containing, say, 20 images, we ran a total of 200 experiments (10 experiments per image). Selecting the best result for each image yields a total of 20 results (one result per image), for which the average is found. Similarly, we calculated the "average worst" (by averaging the worst results for each image). To calculate "average average" and "average standard deviation", instead of selecting one of the 10 runs on every image, we first calculated the average result (and standard deviation) for each image, and then calculated the corresponding averages over all of the images.

*Nathan Netanyahu is also affiliated with the Center for Automation Research, University of Maryland, College Park, MD 20742 (e-mail: nathan@cfar.umd.edu).

2. Description of figures and tables

Figures 1–3 depict a full solution of puzzles of different sizes. Each figure contains the best chromosome in every generation, until a correct solution was reached. Figure 4 extends Figure 2 of the original paper, illustrating the crossover process, *i.e*. the creation of a child chromosome from two parent chromosomes.

Tables 1–5 summarize our solver results on the datasets supplied and used by Pomeranz *et al*. [2]. The first is a set of 20 432-piece puzzles, which first appeared in Cho *et al*. [1]. Each of the next two sets contains 20 images of 540- and 805-piece puzzles. The last two sets contain only 3 images each. The first set consists of 2,360-piece puzzles and the second of 3,300-piece puzzles.

Figures 5–9 show our solver results for all 20 images of our first benchmark of large images, each containing 5,015 puzzle pieces. Figures 10–14 show our solver results for all 20 images of our second benchmark of large images, each containing 10,375 puzzle pieces. Finally, Figures 15–20 show our solver results for all 20 images of our third benchmark of large images, each containing 22,834 puzzle pieces.

References

[1] T. Cho, S. Avidan, and W. Freeman. A probabilistic image jigsaw puzzle solver. In *the IEEE Conference on Computer Vision and Pattern Recognition*, pages 183–190, 2010. 1, 3

[2] D. Pomeranz, M. Shemesh, and O. Ben-Shahar. A fully automated greedy square jigsaw puzzle solver. In *the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9–16, 2011. 1, 5, 6

(a) Original 5,015 puzzle
accuracy: 1.55%

(b) Generation 1
accuracy: 37.88%

(c) Generation 2
accuracy: 72.69%

(d) Generation 3
accuracy: 96.86%

(e) Final generation (4)
accuracy: 100%

Figure 1: Full solution of a 5,015-piece puzzle. This figure details all best chromosomes achieved in every generation until the solution was reached, including their accuracy according to the neighbor measure. Note how the child figure was discovered and assembled correctly in generation 2, but in the wrong absolute location, and how it was detected and shifted in the following generation.

(a) Original 10,375 puzzle
accuracy: 1.09%

(b) Generation 1
accuracy: 35.26%

(c) Generation 2
accuracy: 64.31%

(d) Generation 3
accuracy: 87.72%

(e) Generation 4
accuracy: 96.56%

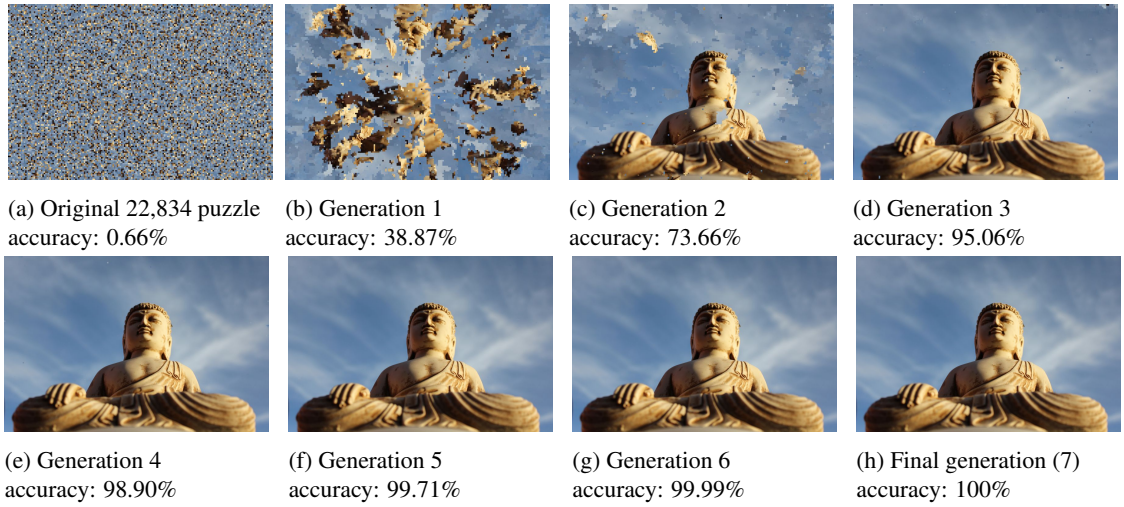
(f) Generation 5
accuracy: 98.50%

(g) Generation 6
accuracy: 99.62%

(h) Generation 7
accuracy: 99.82%

(i) Final generation (8)
accuracy: 100%

Figure 2: Full solution of a 10,375-piece puzzle. This figure details all best chromosomes achieved in every generation until the solution was reached, including their accuracy according to the neighbor measure. Note how the man statue figure was discovered and assembled correctly in generation 2, but in the wrong absolute location, and how it was detected and shifted in the following generation.



(a) Original 22,834 puzzle accuracy: 0.66%
(b) Generation 1 accuracy: 38.87%
(c) Generation 2 accuracy: 73.66%
(d) Generation 3 accuracy: 95.06%
(e) Generation 4 accuracy: 98.90%
(f) Generation 5 accuracy: 99.71%
(g) Generation 6 accuracy: 99.99%
(h) Final generation (7) accuracy: 100%

Figure 3: Full solution of a 22,834-piece puzzle. This figure details all best chromosomes achieved in every generation until the solution was reached, including their accuracy according the neighbor measure.

Image No. (432 pieces)	**Best**	**Worst**	**Average**	**Std. Dev.**
1	88.44	86.62	87.55	0.51
2	85.28	84.18	84.73	0.31
3	100.00	100.00	100.00	0.00
4	69.46	65.09	68.03	1.10
5	100.00	100.00	100.00	0.00
6	98.30	96.72	97.69	0.52
7	100.00	100.00	100.00	0.00
8	100.00	100.00	100.00	0.00
9	100.00	99.64	99.82	0.18
10	97.81	97.81	97.81	0.00
11	97.08	97.08	97.08	0.00
12	99.64	99.03	99.39	0.23
13	91.12	90.15	90.56	0.34
14	99.64	99.64	99.64	0.00
15	96.84	95.86	96.35	0.33
16	100.00	100.00	100.00	0.00
17	99.64	99.64	99.64	0.00
18	100.00	92.82	95.82	3.23
19	100.00	100.00	100.00	0.00
20	100.00	100.00	100.00	0.00
Average	96.16	95.21	95.70	0.34

Table 1: Results of running our GA-based solver 10 times on every image in the 432-piece puzzle set provided by Cho *et al*. [1]. For each image we report the best, worst and average score and the standard deviation. Also, we calculate and report the solver's average performance on the entire set, with respect to the best, worst, and average results obtained over 10 runs per each image.

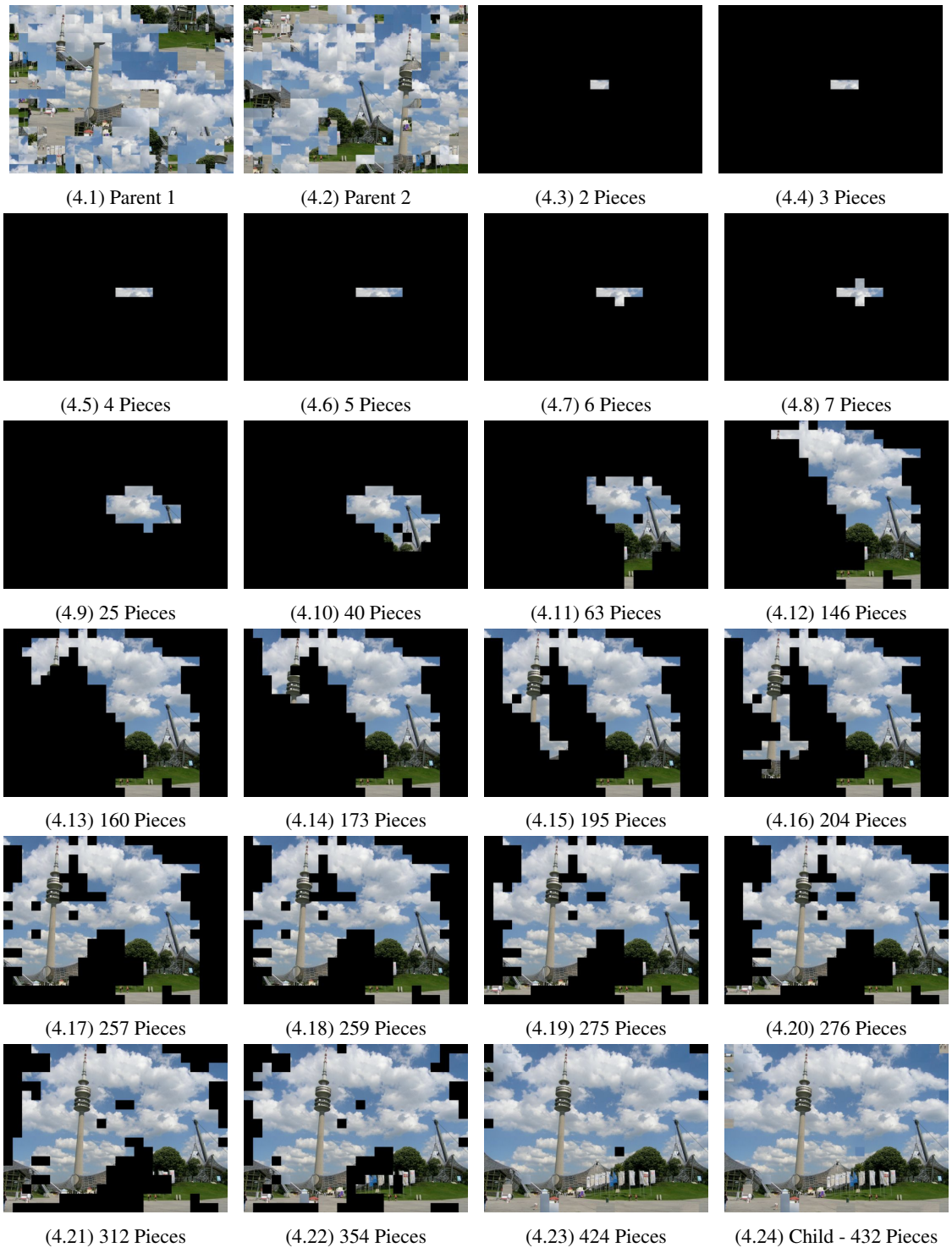


Figure 4: Illustration of crossover operation: Given Parent1 (44.1) and Parent2 (44.2), the above images depict how a kernel of pieces is gradually grown, one piece at a time, until a complete child (44.24) is obtained. Note the detection of parts of the tower in both parents, which are then shifted and merged to the complete tower. Also, note how the entire image is shifted in (44.18), (44.19) and (44.20). This figure is an extension of Figure 2 in the original paper.

Image No. (540 pieces)	Best	Worst	Average	Std. Dev.
1	100.00	100.00	100.00	0.00
2	94.48	93.13	93.78	0.53
3	72.51	68.44	70.15	1.13
4	97.68	96.71	97.30	0.29
5	98.55	98.06	98.36	0.16
6	100.00	100.00	100.00	0.00
7	100.00	100.00	100.00	0.00
8	75.90	71.06	74.18	1.35
9	89.55	82.19	86.84	2.18
10	100.00	100.00	100.00	0.00
11	100.00	100.00	100.00	0.00
12	99.81	99.61	99.70	0.05
13	99.71	99.71	99.71	0.00
14	99.71	99.71	99.71	0.00
15	100.00	100.00	100.00	0.00
16	100.00	100.00	100.00	0.00
17	95.35	90.03	92.72	1.92
18	96.22	95.06	95.52	0.32
19	99.71	99.32	99.67	0.12
20	100.00	100.00	100.00	0.00
Average	95.96	94.65	95.38	0.40

Table 2: Results of running our GA-based solver 10 times on every image in the 540-piece puzzle set provided by Pomeranz *et al*. [2]. For each image we report the best, worst and average score and the standard deviation. Also, we calculate and report the solver's average performance on the entire set, with respect to the best, worst, and average results obtained over 10 runs per each image.

Image No. (805 pieces)	Best	Worst	Average	Std. Dev.
1	87.76	85.76	86.76	0.65
2	91.49	90.46	90.82	0.36
3	94.78	93.49	94.22	0.39
4	90.66	89.30	90.03	0.44
5	100.00	100.00	100.00	0.00
6	100.00	97.49	99.34	1.02
7	100.00	100.00	100.00	0.00
8	100.00	99.74	99.84	0.08
9	98.32	97.49	97.89	0.32
10	100.00	100.00	100.00	0.00
11	99.81	98.78	99.29	0.34
12	80.73	75.71	78.84	1.71
13	100.00	100.00	100.00	0.00
14	100.00	100.00	100.00	0.00
15	100.00	100.00	100.00	0.00
16	92.33	91.43	91.85	0.30
17	100.00	100.00	100.00	0.00
18	90.40	89.24	89.64	0.33
19	100.00	100.00	100.00	0.00
20	98.90	98.20	98.50	0.25
Average	96.26	95.35	95.85	0.31

Table 3: Results of running our GA-based solver 10 times on every image in the 805-piece puzzle set provided by Pomeranz *et al*. [2]. For each image we report the best, worst and average score and the standard deviation. Also, we calculate and report the solver's average performance on the entire set, with respect to the best, worst, and average results obtained over 10 runs per each image.

Image No. (2,360 pieces)	Best	Worst	Average	Std. Dev.
1	100.00	100.00	100.00	0.00
2	80.44	78.53	79.16	0.52
3	86.15	84.03	84.86	0.60
Average	88.86	87.52	88.01	0.38

Table 4: Results of running our GA-based solver 10 times on every image in the 2,360-piece puzzle set provided by Pomeranz *et al*. [2]. For each image we report the best, worst and average score and the standard deviation. Also, we calculate and report the solver's average performance on the entire set, with respect to the best, worst, and average results obtained over 10 runs per each image.

Image No. (3,300 pieces)	Best	Worst	Average	Std. Dev.
1	100.00	100.00	100.00	0.00
2	80.44	78.53	79.16	0.52
3	86.15	84.03	84.86	0.60
Average	88.86	87.52	88.01	0.38

Table 5: Results of running our GA-based solver 10 times on every image in the 3,300-piece puzzle set provided by Pomeranz *et al*. [2]. For each image we report the best, worst and average score and the standard deviation. Also, we calculate and report the solver's average performance on the entire set, with respect to the best, worst, and average results obtained over 10 runs per each image.

Figure 5: Results of running our GA-based solver on every image in the 5,015-piece puzzle set we created. For each image i, panels $i-a$, $i-b$ and $i-c$ show the best chromosome created in the first, second and last generations, respectively. Following are the best, worst and average results and standard deviation, under the neighbor comparison, obtained from 10 runs:
Image 1: best: 99.55%, worst: 99.02%, average: 99.18%, standard deviation: 0.14%
Image 2: best: 100.00%, worst: 99.89%, average: 99.92%, standard deviation: 0.03%
Image 3: best: 100.00%, worst: 100.00%, average: 100.00%, standard deviation: 100.00%

Figure 6: Results of running our GA-based solver on every image in the 5,015-piece puzzle set we created. For each image i, panels $i-a$, $i-b$ and $i-c$ show the best chromosome created in the first, second and last generations, respectively. Following are the best, worst and average results and standard deviation, under the neighbor comparison, obtained from 10 runs:
Image 4: best: 86.17%, worst: 85.27%, average: 85.73%, standard deviation: 0.25%
Image 5: best: 71.80%, worst: 70.56%, average: 71.19%, standard deviation: 0.33%
Image 6: best: 100.00%, worst: 100.00%, average: 100.00%, standard deviation: 100.00%

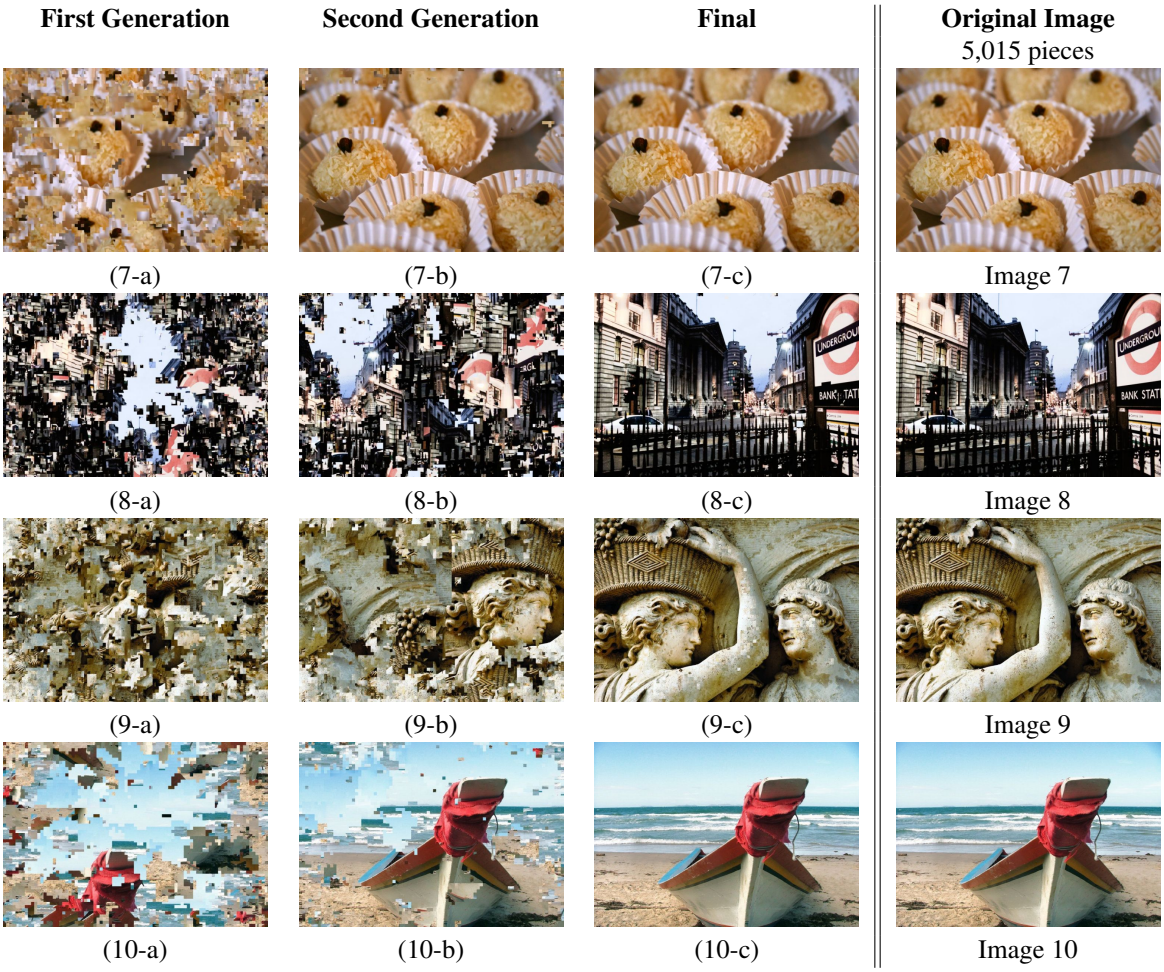


Figure 7: Results of running our GA-based solver on every image in the 5,015-piece puzzle set we created. For each image i, panels $i-a$, $i-b$ and $i-c$ show the best chromosome created in the first, second and last generations, respectively. Following are the best, worst and average results and standard deviation, under the neighbor comparison, obtained from 10 runs:
Image 7: best: 100.00%, worst: 100.00%, average: 100.00%, standard deviation: 0.00%
Image 8: best: 93.62%, worst: 91.56%, average: 92.39%, standard deviation: 0.55%
Image 9: best: 92.46%, worst: 91.82%, average: 92.18%, standard deviation: 0.21%
Image 10: best: 100.00%, worst: 99.96%, average: 99.98%, standard deviation: 0.02%

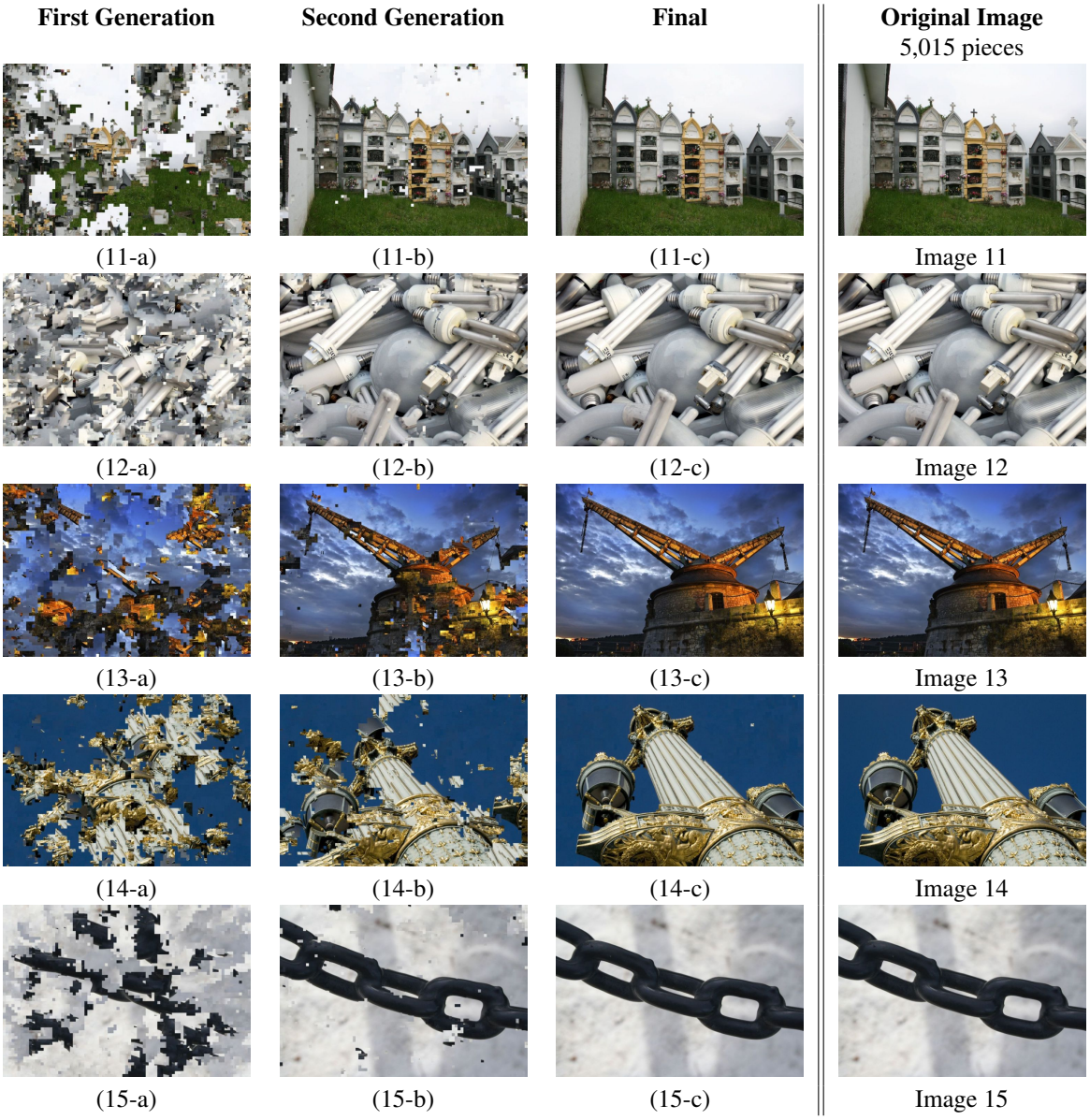


Figure 8: Results of running our GA-based solver on every image in the 5,015-piece puzzle set we created. For each image i, panels $i-a$, $i-b$ and $i-c$ show the best chromosome created in the first, second and last generations, respectively. Following are the best, worst and average results and standard deviation, under the neighbor comparison, obtained from 10 runs:
Image 11: best: 96.01%, worst: 95.36%, average: 95.77%, standard deviation: 0.20%
Image 12: best: 100.00%, worst: 100.00%, average: 100.00%, standard deviation: 0.00%
Image 13: best: 100.00%, worst: 99.95%, average: 99.97%, standard deviation: 0.01%
Image 14: best: 65.65%, worst: 64.57%, average: 65.25%, standard deviation: 0.32%
Image 15: best: best: 100.00%, worst: 100.00%, average: 100.00%, standard deviation: 0.00%

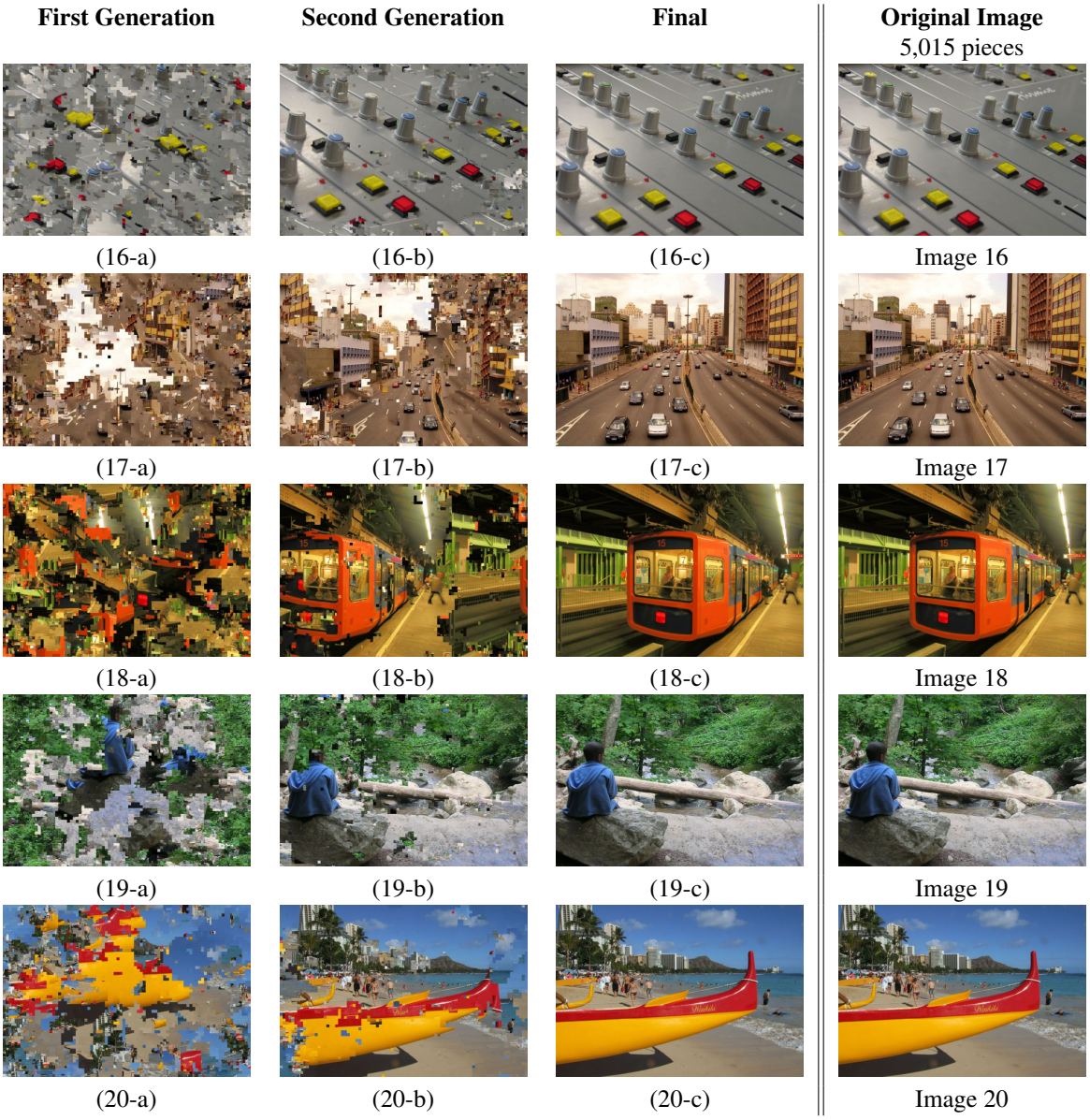


Figure 9: Results of running our GA-based solver on every image in the 5,015-piece puzzle set we created. For each image i, panels $i-a$, $i-b$ and $i-c$ show the best chromosome created in the first, second and last generations, respectively. Following are the best, worst and average results and standard deviation, under the neighbor comparison, obtained from 10 runs:
Image 16: best: 100.00%, worst: 100.00%, average: 100.00%, standard deviation: 0.00%
Image 17: best: 99.70%, worst: 99.58%, average: 99.64%, standard deviation: 0.04%
Image 18: best: 100.00%, worst: 99.97%, average: 99.97%, standard deviation: 0.01%
Image 19: best: best: 100.00%, worst: 100.00%, average: 100.00%, standard deviation: 0.00%
Image 20: best: best: 99.97%, worst: 99.90%, average: 99.95%, standard deviation: 0.02%

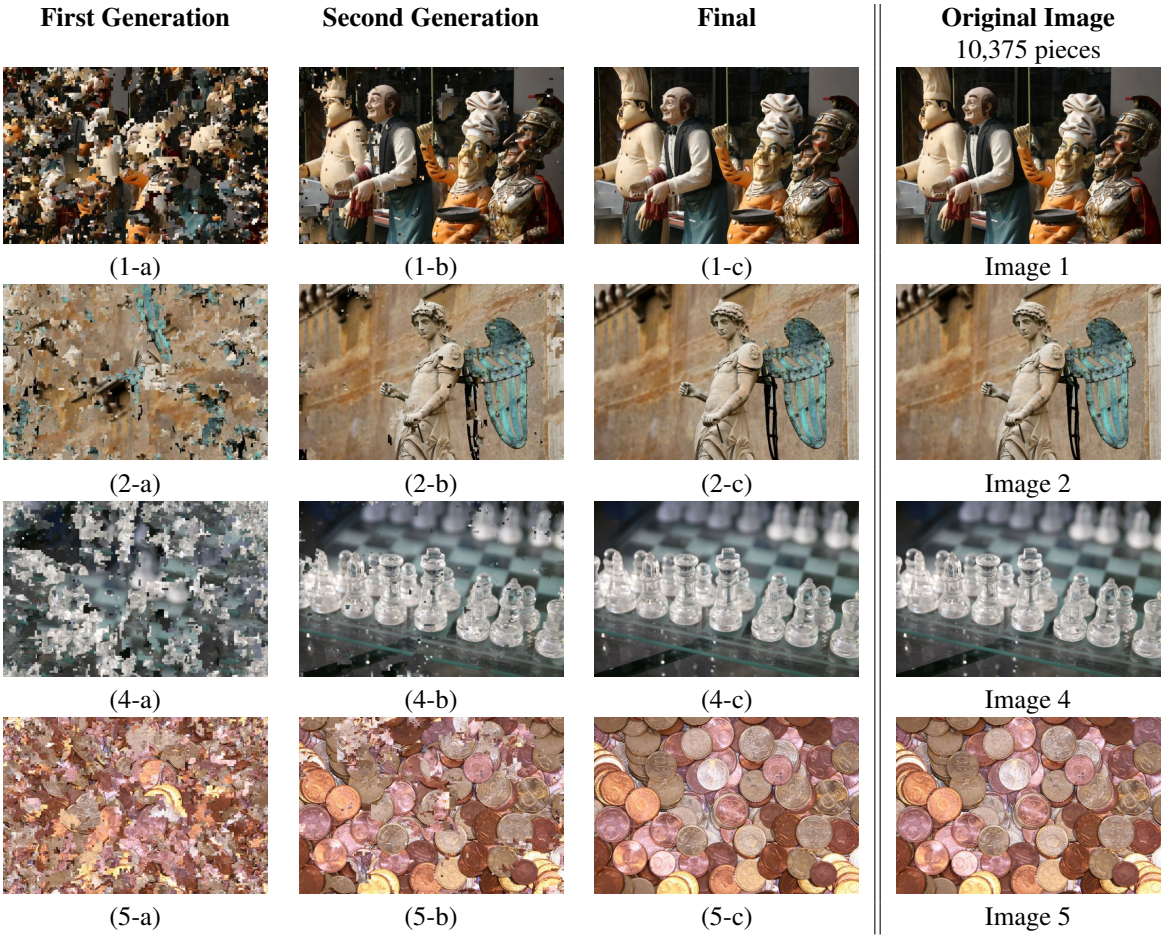


Figure 10: Results of running our GA-based solver on every image in the 10,375-piece puzzle set we created. For each image i, panels $i-a$, $i-b$ and $i-c$ show the best chromosome created in the first, second and last generations, respectively. Following are the best, worst and average results and standard deviation, under the neighbor comparison, obtained from 10 runs:
image 1: best: 99.89, worst: 99.73, average: 99.82, standard deviation: 0.05
image 2: best: 99.98, worst: 99.93, average: 99.95, standard deviation: 0.01
image 4: best: 100.00, worst: 99.99, average: 100.00, standard deviation: 0.00
image 5: best: 100.00, worst: 100.00, average: 100.00, standard deviation: 0.00

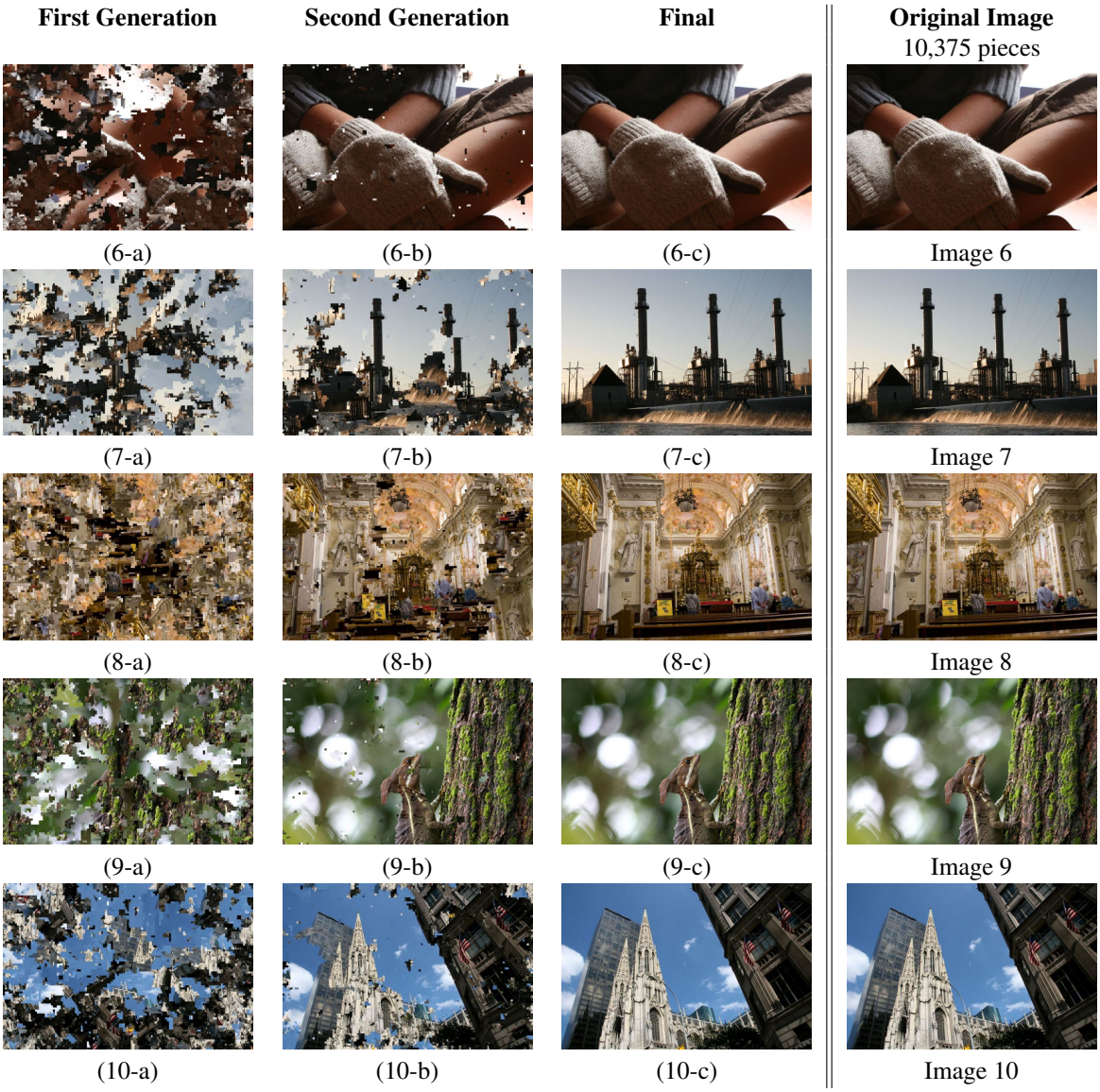


Figure 11: Results of running our GA-based solver on every image in the 10,375-piece puzzle set we created. For each image i, panels $i-a$, $i-b$ and $i-c$ show the best chromosome created in the first, second and last generations, respectively. Following are the best, worst and average results and standard deviation, under the neighbor comparison, obtained from 10 runs:
image 6: best: 97.46, worst: 97.39, average: 97.42, standard deviation: 0.02
image 7: best: 99.78, worst: 99.43, average: 99.68, standard deviation: 0.09
image 8: best: 99.79, worst: 99.72, average: 99.75, standard deviation: 0.02
image 9: best: 99.15, worst: 99.06, average: 99.10, standard deviation: 0.03
image 10: best: 98.84, worst: 98.56, average: 98.66, standard deviation: 0.07

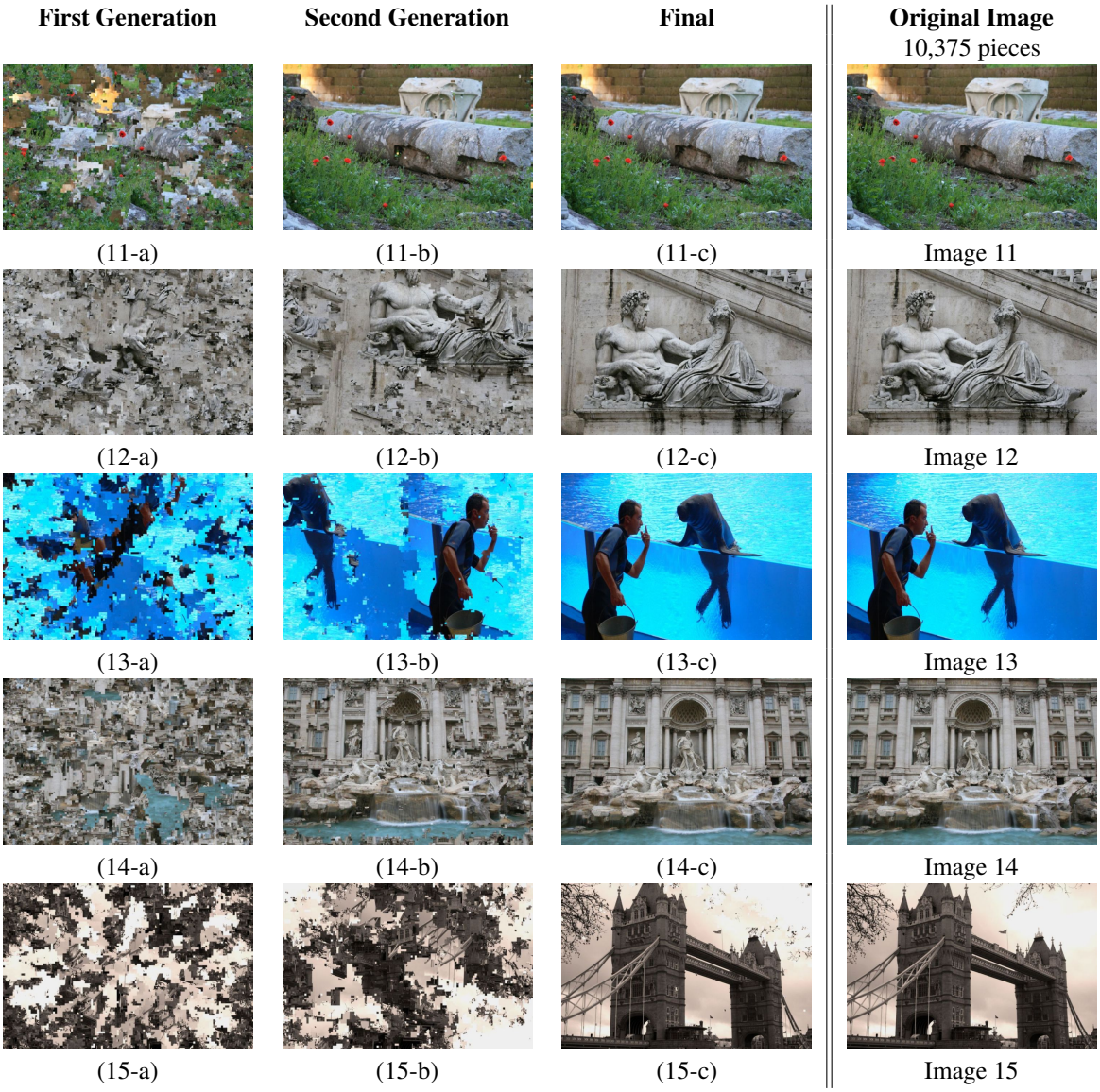


Figure 12: Results of running our GA-based solver on every image in the 10,375-piece puzzle set we created. For each image i, panels $i-a$, $i-b$ and $i-c$ show the best chromosome created in the first, second and last generations, respectively. Following are the best, worst and average results and standard deviation, under the neighbor comparison, obtained from 10 runs:
image 11: best: 100.00, worst: 100.00, average: 100.00, standard deviation: 0.00
image 12: best: 100.00, worst: 100.00, average: 100.00, standard deviation: 0.00
image 13: best: 99.39, worst: 99.14, average: 99.30, standard deviation: 0.08
image 14: best: 100.00, worst: 99.98, average: 100.00, standard deviation: 0.01
image 15: best: 85.00, worst: 84.37, average: 84.73, standard deviation: 0.22

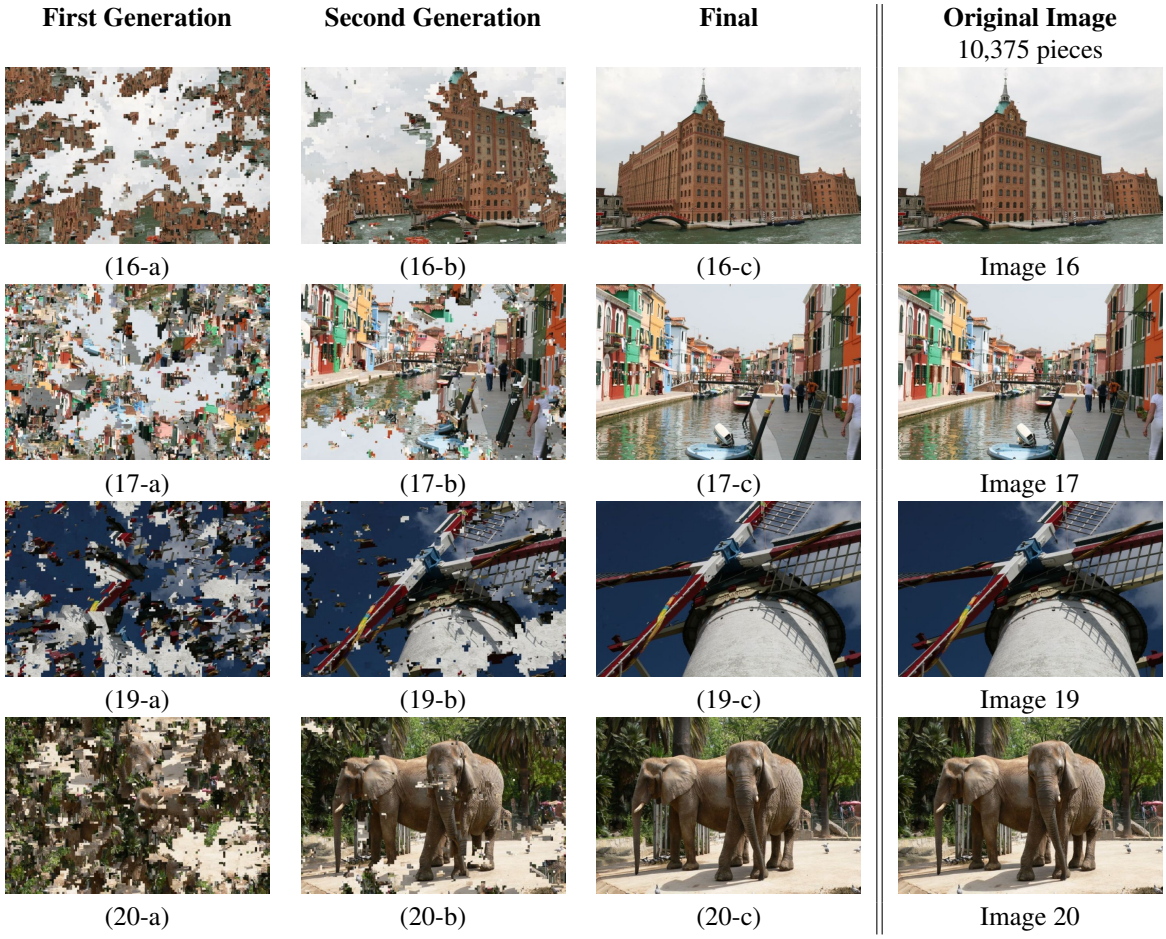


Figure 13: Results of running our GA-based solver on every image in the 10,375-piece puzzle set we created. For each image i, panels $i-a$, $i-b$ and $i-c$ show the best chromosome created in the first, second and last generations, respectively. Following are the best, worst and average results and standard deviation, under the neighbor comparison, obtained from 10 runs:
image 16: best: 97.39, worst: 96.46, average: 97.12, standard deviation: 0.25
image 17: best: 96.90, worst: 95.82, average: 96.44, standard deviation: 0.33
image 19: best: 95.90, worst: 94.76, average: 95.42, standard deviation: 0.42
image 20: best: 100.00, worst: 99.99, average: 99.99, standard deviation: 0.01

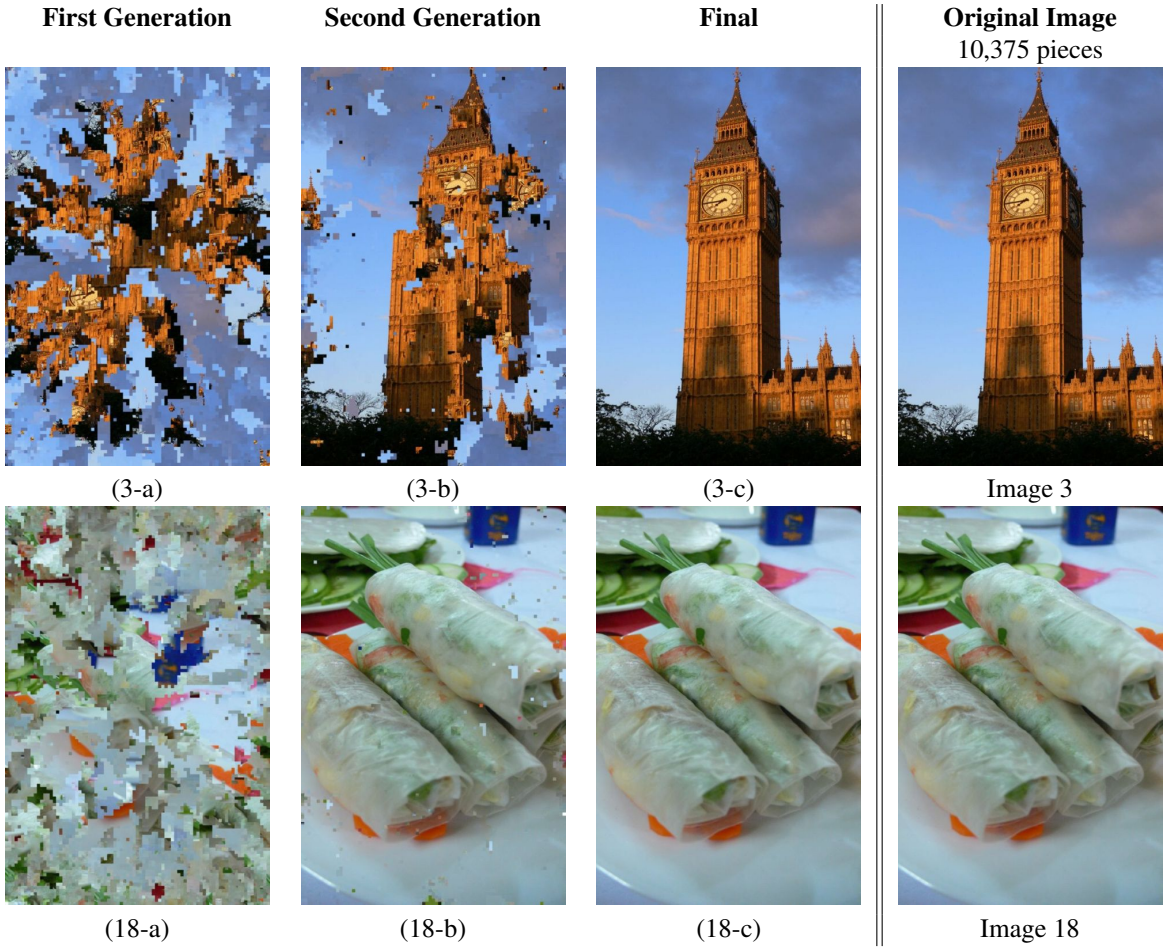


Figure 14: Results of running our GA-based solver on every image in the 10,375-piece puzzle set we created. For each image i, panels $i-a$, $i-b$ and $i-c$ show the best chromosome created in the first, second and last generations, respectively. Following are the best, worst and average results and standard deviation, under the neighbor comparison, obtained from 10 runs:
image 3: best: 99.87, worst: 99.73, average: 99.79, standard deviation: 0.05
image 18: best: 100.00, worst: 100.00, average: 100.00, standard deviation: 0.00

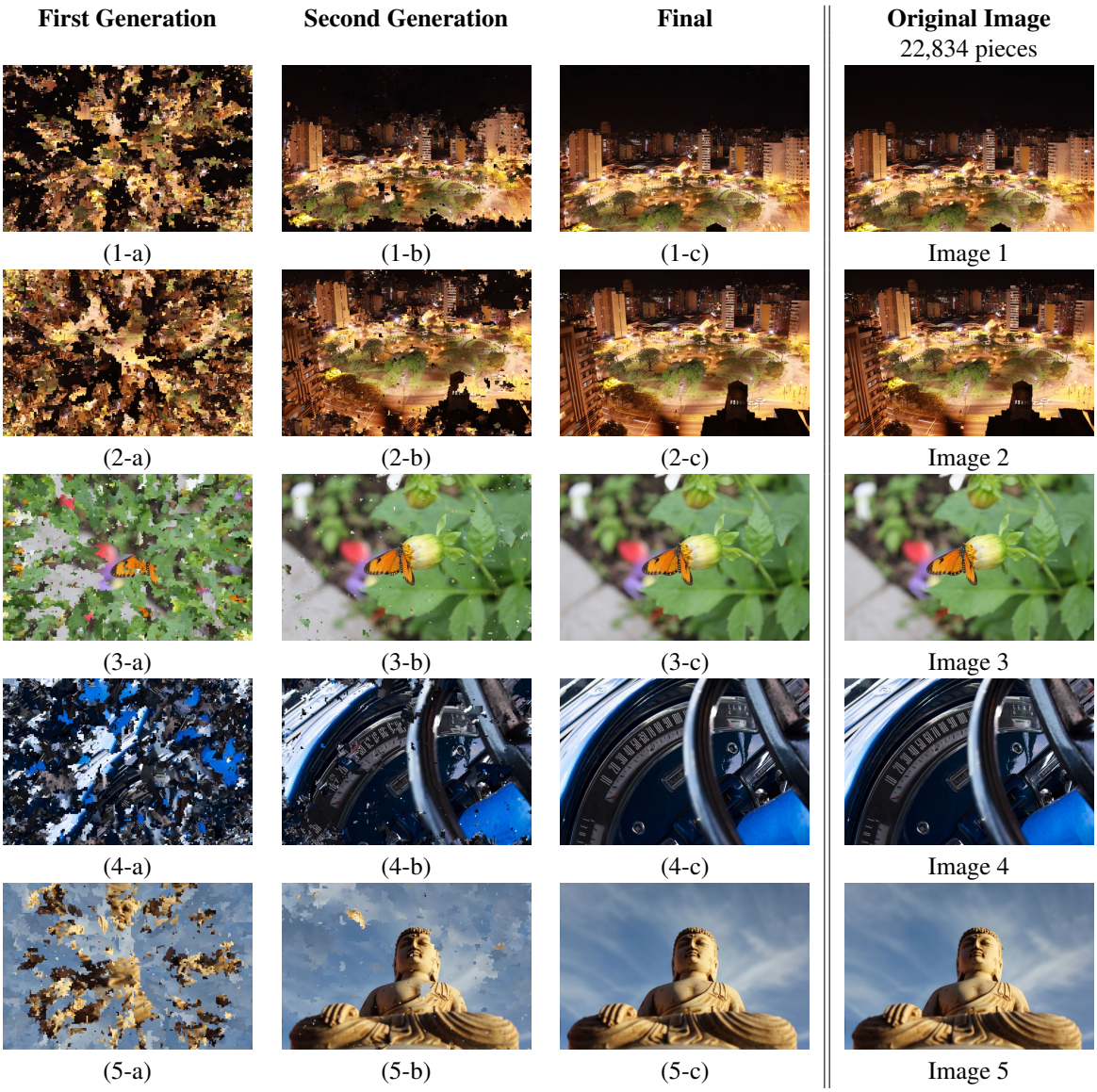


Figure 15: Results of running our GA-based solver on every image in the 22,834-piece puzzle set we created. For each image i, panels $i-a$, $i-b$ and $i-c$ show the best chromosome created in the first, second and last generations, respectively. Following are the best, worst and average results and standard deviation, under the neighbor comparison, obtained from 2 runs:
image 1: best: 94.56, worst: 94.39, average: 94.48, standard deviation: 0.09
image 2: best: 99.12, worst: 99.07, average: 99.10, standard deviation: 0.03
image 3: best: 99.91, worst: 99.87, average: 99.89, standard deviation: 0.02
image 4: best: 99.61, worst: 99.57, average: 99.59, standard deviation: 0.02
image 5: best: 100.00, worst: 100.00, average: 100.00, standard deviation: 0.00

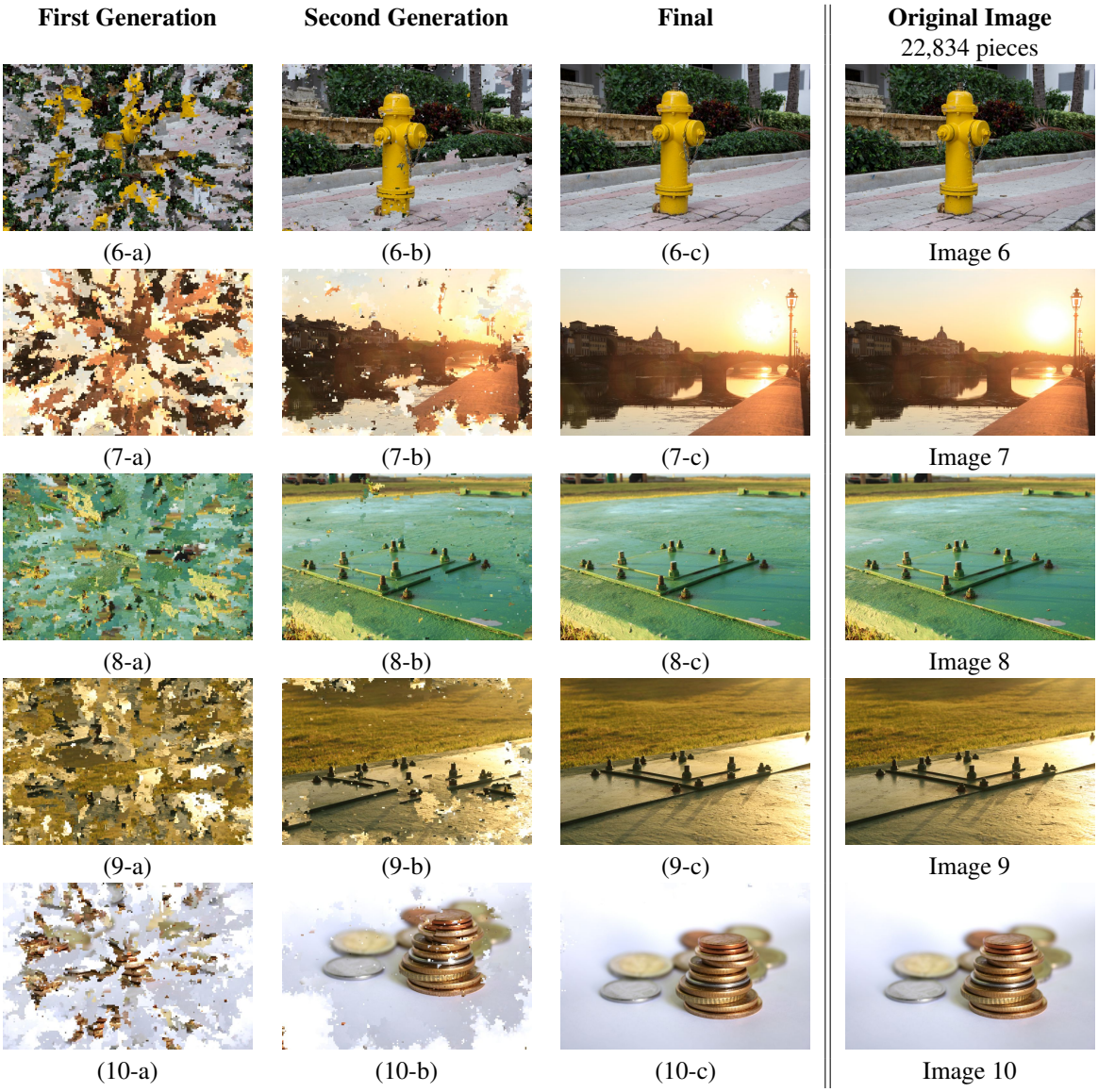


Figure 16: Results of running our GA-based solver on every image in the 22,834-piece puzzle set we created. For each image i, panels $i-a$, $i-b$ and $i-c$ show the best chromosome created in the first, second and last generations, respectively. Following are the best, worst and average results and standard deviation, under the neighbor comparison, obtained from 2 runs:
image 6: best: 99.76, worst: 99.75, average: 99.76, standard deviation: 0.00
image 7: best: 95.05, worst: 94.76, average: 94.91, standard deviation: 0.14
image 8: best: 100.00, worst: 100.00, average: 100.00, standard deviation: 0.00
image 9: best: 99.25, worst: 99.24, average: 99.24, standard deviation: 0.01
image 10: best: 82.41, worst: 82.20, average: 82.30, standard deviation: 0.10

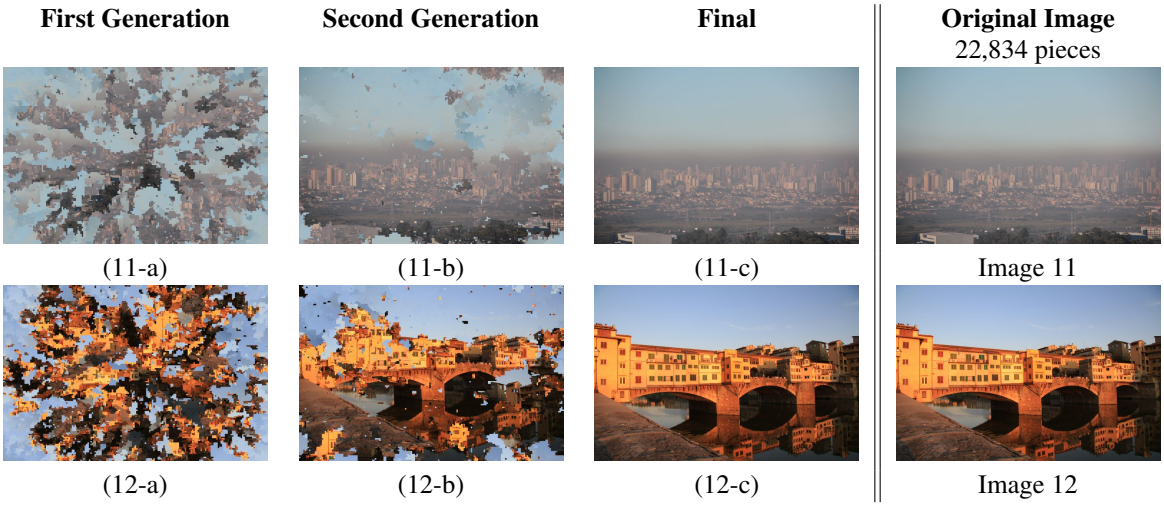


Figure 17: Results of running our GA-based solver on every image in the 22,834-piece puzzle set we created. For each image i, panels $i-a$, $i-b$ and $i-c$ show the best chromosome created in the first, second and last generations, respectively. Following are the best, worst and average results and standard deviation, under the neighbor comparison, obtained from 2 runs:
image 11: best: 99.86, worst: 99.83, average: 99.85, standard deviation: 0.01
image 12: best: 99.99, worst: 99.99, average: 99.99, standard deviation: 0.00

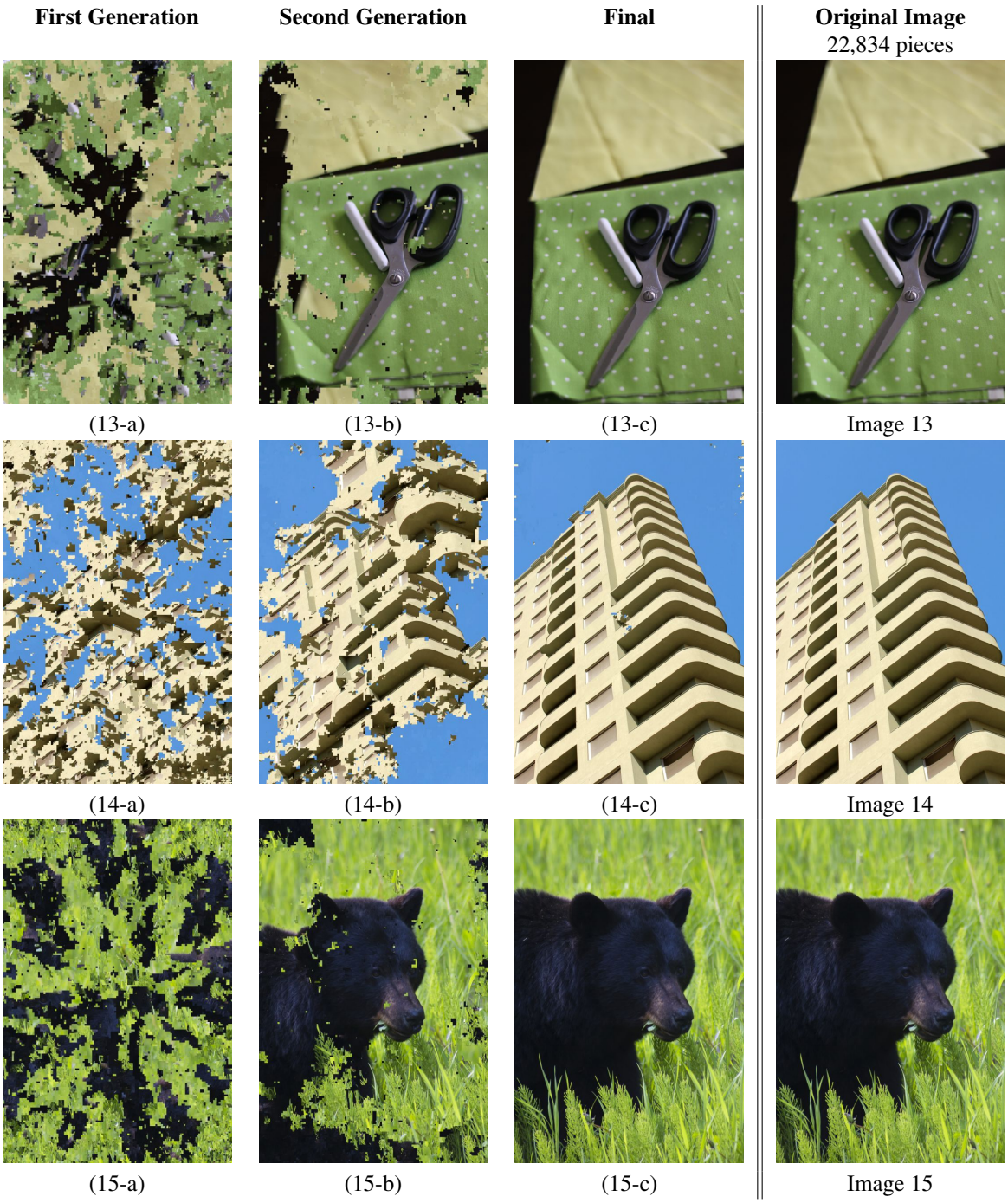


Figure 18: Results of running our GA-based solver on every image in the 22,834-piece puzzle set we created. For each image i, panels $i-a$, $i-b$ and $i-c$ show the best chromosome created in the first, second and last generations, respectively. Following are the best, worst and average results and standard deviation, under the neighbor comparison, obtained from 2 runs:
image 13: best: 100.00, worst: 100.00, average: 100.00, standard deviation: 0.00
image 14: best: 90.21, worst: 90.09, average: 90.15, standard deviation: 0.06
image 15: best: 99.38, worst: 99.38, average: 99.38, standard deviation: 0.00

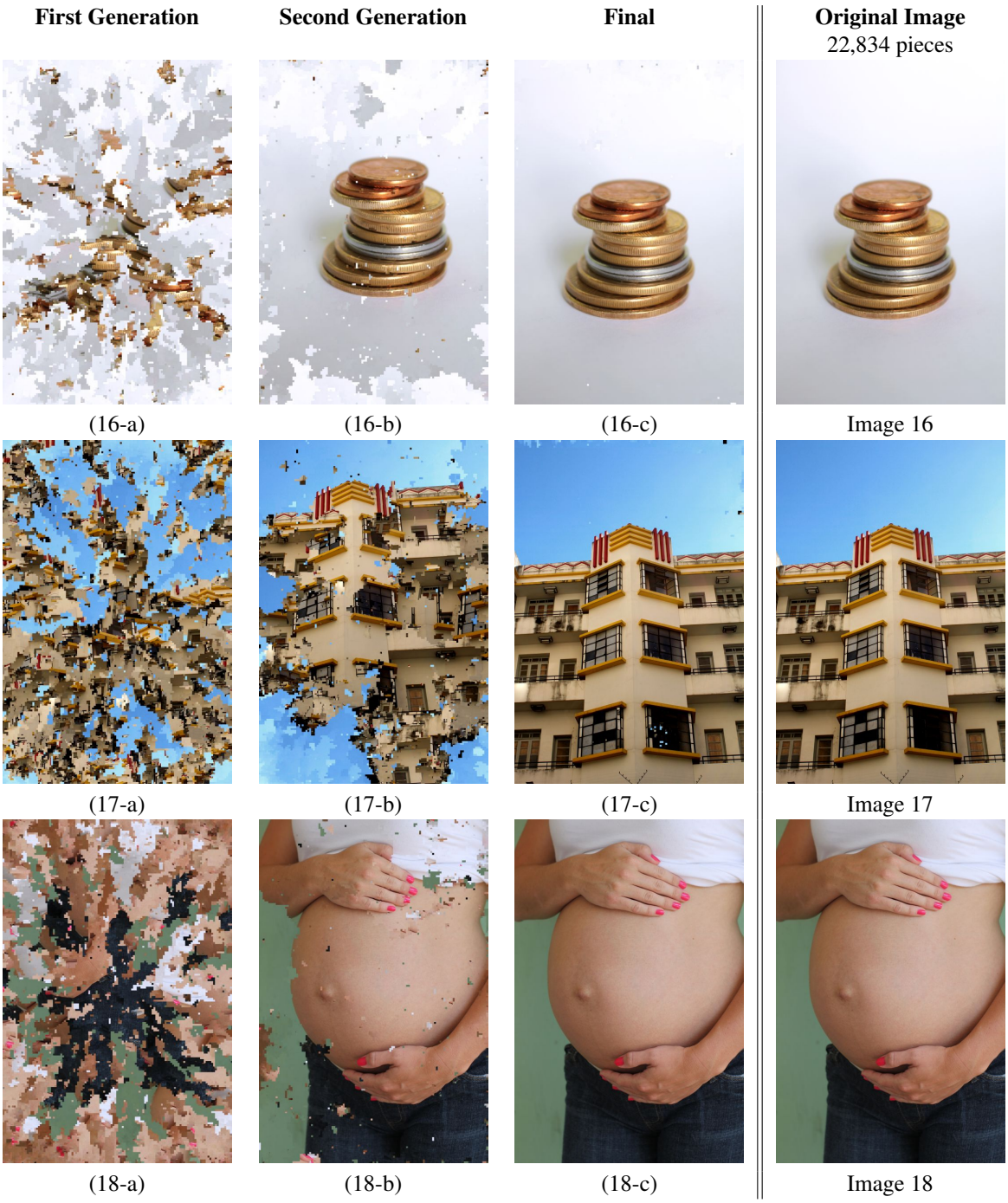


Figure 19: Results of running our GA-based solver on every image in the 22,834-piece puzzle set we created. For each image i, panels $i-a$, $i-b$ and $i-c$ show the best chromosome created in the first, second and last generations, respectively. Following are the best, worst and average results and standard deviation, under the neighbor comparison, obtained from 2 runs:
image 16: best: 89.71, worst: 89.27, average: 89.49, standard deviation: 0.22
image 17: best: 95.49, worst: 95.36, average: 95.42, standard deviation: 0.07
image 18: best: 100.00, worst: 100.00, average: 100.00, standard deviation: 0.00

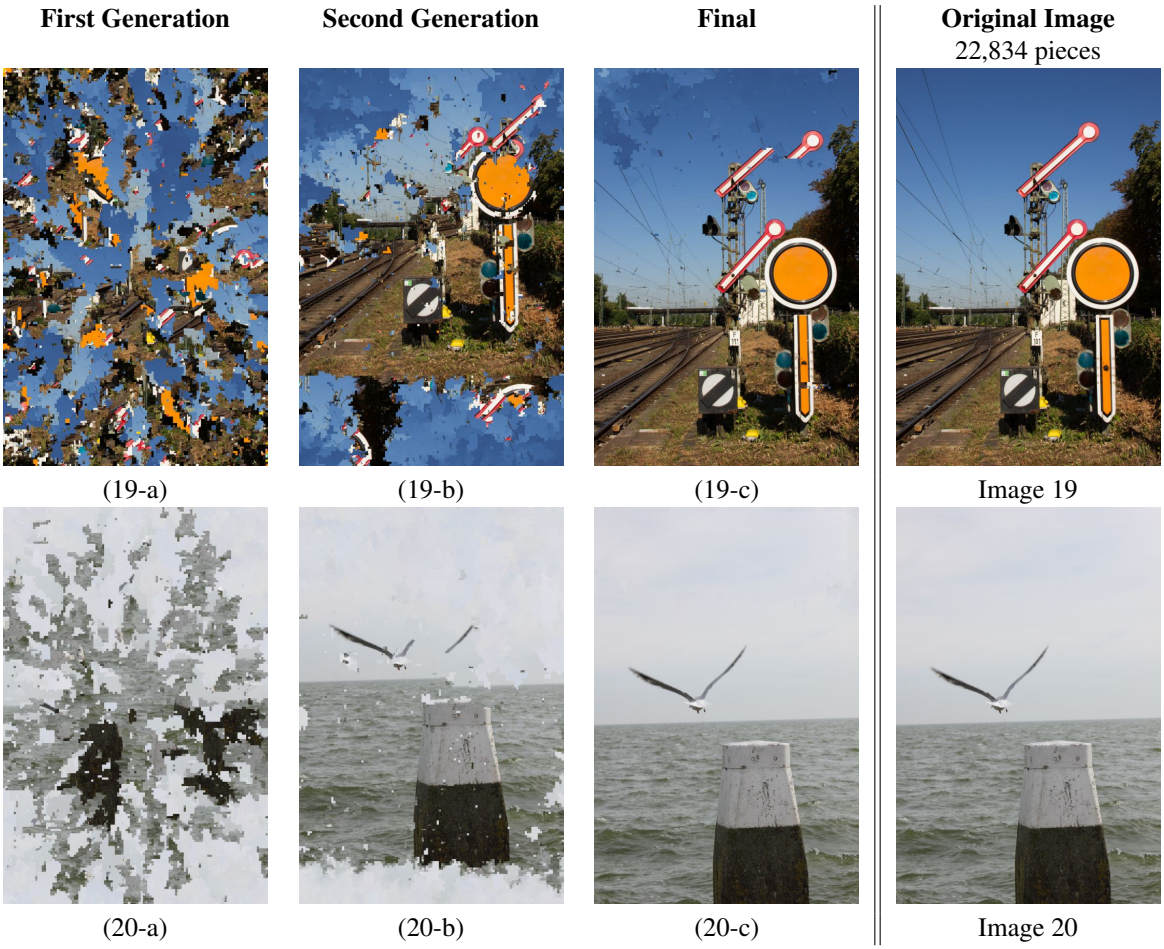


Figure 20: Results of running our GA-based solver on every image in the 22,834-piece puzzle set we created. For each image i, panels $i-a$, $i-b$ and $i-c$ show the best chromosome created in the first, second and last generations, respectively. Following are the best, worst and average results and standard deviation, under the neighbor comparison, obtained from 2 runs:
image 19: best: 87.13, worst: 86.95, average: 87.04, standard deviation: 0.09
image 20: best: 94.08, worst: 93.29, average: 93.69, standard deviation: 0.39