Table 1: Scores for 2D problems. Mean scores and standard deviations are shown over 7 training instances in the left three columns. The right three columns show results of statistical tests in binary comparisons to p=0.5 confidence. A green check is marked when GAN-MC significantly outperforms its counterpart. Red X and orange check marks denote cases where GAN or GAN-DO significantly outperform their counterpart. A gray dash indicates that neither model outperforms the other to the desired statistical certainty.

	GAN	**GAN-DO**	**GAN-MC**	**GAN-DO ✓ GAN ✗**	**GAN-MC ✓ GAN ✗**	**GAN-MC ✓ GAN-DO ✓**
			Problem 1			
Invalidity (%) ↓	7.41±3.67	**0.11±0.05**	0.14±0.06	✓	✓	–
MMD (E-3) ↓	7.21±1.94	3.17±0.12	**2.66±0.04**	✓	✓	✓
F1 ↑	.392±0.108	0.831±0.25	**0.850±0.011**	✓	✓	✓
			Problem 2			
Invalidity (%) ↓	3.27±0.66	**0.24±0.07**	0.56±0.23	✓	✓	✓
MMD (E-3) ↓	2.57±0.03%	2.53±0.01%	**2.43±0.01%**	✓	✓	✓
F1 ↑	0.829±0.012	0.863±0.011	**0.876±0.006**	✓	✓	✓

Table 2: Remake of Table 2 in original paper with GAN-MC, showing invalidity metric. GAN scores are also re-evaluated, showing less anomalous results. Standard deviations over the 4 runs are included. Any GAN-MC score that is significant to $p < 0.05$ over the corresponding GAN score (same quantity of positive data) is bolded. **Lower is better.**

(a) Models

	Negative Samples
GAN	0
GAN-MC	$1K$
GAN-MC	$4K$
GAN-MC	$16K$

(b) Problem 1

Positive Samples $1K$	Positive Samples $4K$	Positive Samples $16K$
10.7% ± 5.3%	11.6% ± 2.9%	11.9% ± 1.6%
2.0% ± 2.3%	**0.7% ± 0.2%**	**1.5% ± 0.8%**
0.5% ± 0.3%	**0.5% ± 0.3%**	**0.5% ± 0.2%**
0.5% ± 0.3%	**0.7% ± 0.5%**	**0.6% ± 0.3%**

(c) Problem 2

Positive Samples $1K$	Positive Samples $4K$	Positive Samples $16K$
6.0% ± 1.5%	3.2% ± 1.3%	3.3% ± 0.7%
2.3% ± 0.7%	1.9% ± 0.7%	**1.6% ± 0.7%**
0.6% ± 0.2%	**0.7% ± 0.3%**	**0.7% ± 0.2%**
0.5% ± 0.1%	**0.2% ± 0.1%**	**0.2% ± 0.1%**

Table 3: Invalidity Rates (%) with Standard Deviations for engineering datasets. Lower scores are better. Problems are sorted by GAN validity rate (harder problems at bottom). Mean scores and standard deviations are shown over 7 training instances in the left three columns. The right three columns show results of statistical tests in binary comparisons to p=0.5 confidence. A green check is marked when GAN-MC significantly outperforms its counterpart. Red X and orange check marks denote cases where GAN or GAN-DO significantly outperform their counterpart. A gray dash indicates that neither model outperforms the other to the desired statistical certainty.

	GAN	GAN-DO	GAN-MC	GAN-DO ✓ GAN ✗	GAN-MC ✓ GAN ✗	GAN-MC ✓ GAN-DO ✓
Three-Bar Truss	0.32±0.51%	**0.00±0.00**%	0.34±0.49%	–	–	✓
Gearbox	0.33±0.09%	**0.02±0.02**%	0.07±0.05%	✓	✓	✓
Concrete Beam	1.03±0.97%	**0.16±0.14**%	1.14±0.46%	✓	–	✓
Pressure Vessel	1.30±0.32%	**0.11±0.07**%	0.95±0.29%	✓	✓	✓
Comp. Spring	1.49±1.00%	**0.77±0.67**%	1.06±0.58%	–	–	–
Ashby Chart	1.54±0.97%	1.12±0.39%	**0.63±0.17**%	–	✓	✓
Welded Beam	1.74±0.89%	0.67±0.39%	**0.53±0.14**%	✓	✓	–
Cantilever Beam	4.16±0.87%	**2.51±0.79**%	3.00±0.62%	✓	✓	–
Bike Frame	4.77±1.21%	**2.85±0.63**%	6.51±3.04%	✓	–	✓
Car Impact	4.78±0.55%	**1.92±0.48**%	3.84±0.78%	✓	✓	✓
Heat Exchanger	5.35±1.00%	3.77±0.70%	**3.68±0.82**%	✓	✓	–
Ship Hull	93.97±0.64%	93.54±0.97%	**92.05±2.31**%	–	✓	–

Table 4: F1 scores with Standard Deviations for engineering datasets. Higher scores are better. Problems are sorted by GAN F1 score (harder problems at bottom). Mean scores and standard deviations are shown over 7 training instances in the left three columns. The right three columns show results of statistical tests in binary comparisons to p=0.5 confidence. A green check is marked when GAN-MC significantly outperforms its counterpart. Red X and orange check marks denote cases where GAN or GAN-DO significantly outperform their counterpart. A gray dash indicates that neither model outperforms the other to the desired statistical certainty.

	GAN	GAN-DO	GAN-MC	GAN-DO ✓ GAN ✗	GAN-MC ✓ GAN ✗	GAN-MC ✓ GAN-DO ✓
Compression Spring	0.960±0.003	0.956±0.004	**0.962±0.005**	✗	–	✓
Ashby Chart	0.959±0.007	**0.960±0.005**	0.922±0.014	–	✗	✓
Concrete Beam	**0.957±0.002**	0.954±0.004	0.956±0.005	–	–	–
Welded Beam	**0.955±0.006**	0.936±0.013	0.850±0.025	✗	✗	✓
Three-Bar Truss	0.938±0.022	0.948±0.012	**0.957±0.005**	–	✓	✓
Pressure Vessel	**0.947±0.012**	0.944±0.013	0.932±0.013	–	✗	–
Gearbox	**0.899±0.023**	0.872±0.021	0.891±0.018	✗	–	✓
Car Impact	0.883±0.017	0.844±0.041	**0.893±0.010**	✗	–	✓
Heat Exchanger	**0.876±0.035**	0.869±0.023	0.867±0.021	–	–	–
Cantilever Beam	0.845±0.038	0.818±0.027	**0.875±0.018**	–	✓	✓
Ship Hull	**0.769±0.082**	0.708±0.273	0.713±0.248	–	–	–
Bike Frame	0.681±0.030	0.684±0.025	**0.741±0.010**	–	✓	✓

Table 5: Updated topology optimization experiments averaged over 1-2 runs (for now). We have benchmarked and added GAN-DO. Instead of reporting average number of floating pixels, we instead report the percentage of the image that is floating, which we call constraint violation magnitude.

	Invalidity (%) ↓	Violation Magnitude (%) ↓
GAN	36.3	0.254
GAN-DO (Synthetic negative data)	22.2	0.329
GAN-MC (Synthetic negative data)	**18.9**	**0.145**
GAN-DO (Rejected negative data)	**12.6**	**0.085**
GAN-MC (Rejected negative data)	16.0	0.118

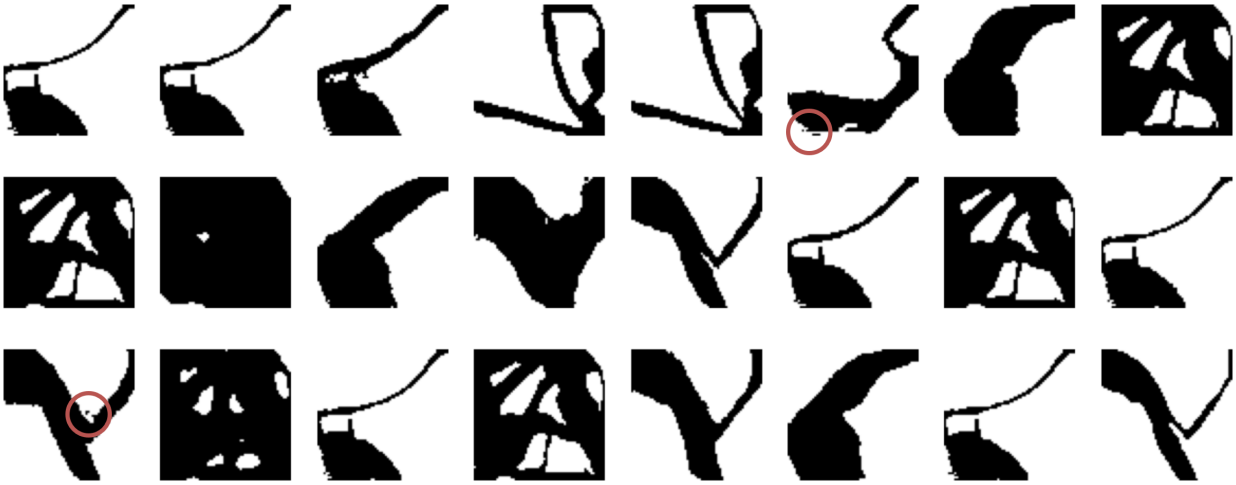

Figure 1: Randomly-selected topologies generated by GAN-DO trained on rejected negative data with constraint violations annotated.

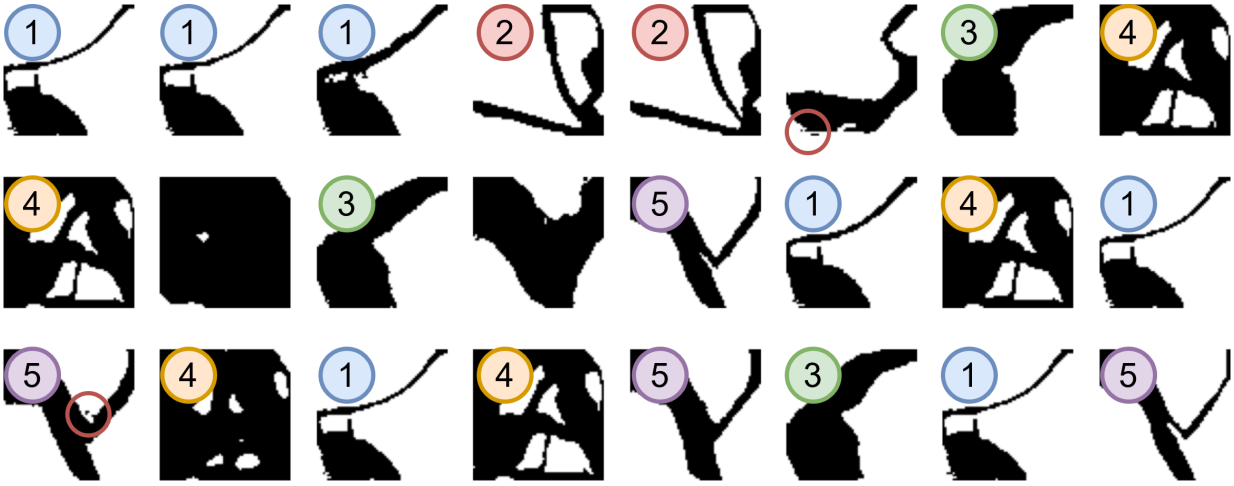

Figure 2: Randomly-selected topologies generated by GAN-DO manually annotated into groups.

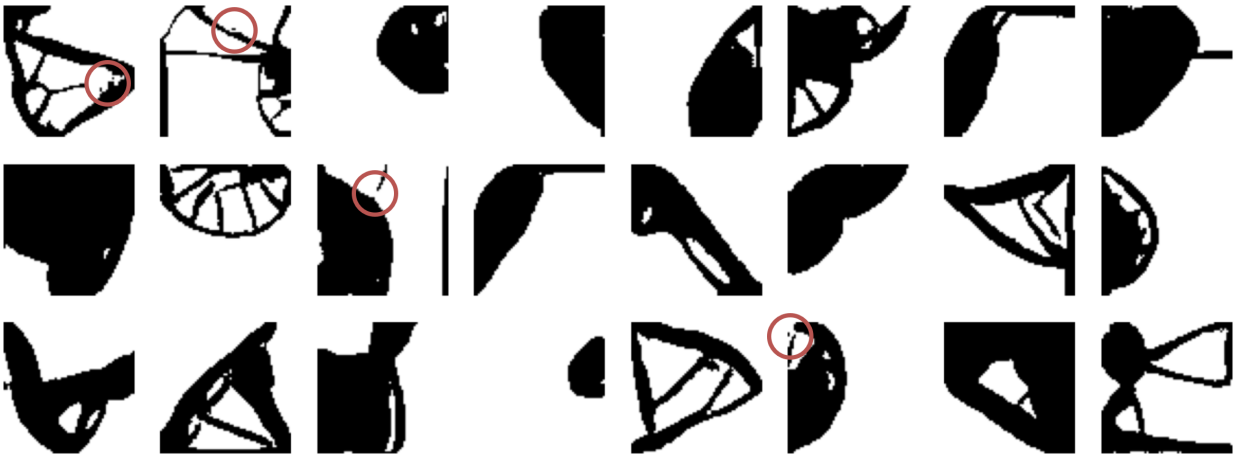

Figure 3: Randomly-selected topologies generated by GAN-MC trained on rejected negative data with constraint violations annotated (taken from submitted paper draft). GAN-MC topologies do not have a visible issue with diversity.