

Geometric Morphometric Data Augmentation using Generative Computational Learning Algorithms

Lloyd A. Courtenay & Diego González-Aguilera

Supplementary Files and Tables

INDEX

Table S1 – Page 2

Table S2 – Page 3

Table S3 – Page 4

Table S4 – Page 5

Table S5 – Page 6

Table S6 – Page 7

	Original Data		LSGAN		WGAN		WGAN-GP	
	PC1*	PC2	PC1*	PC2	PC1*	PC2*	PC1*	PC2
Shapiro w	0.95	0.97	0.94	0.98	0.91	0.93	0.96	0.96
Shapiro p	0.02	0.15	0.005	0.53	4.0e-04	0.002	0.05	0.08
Central Tendency [‡]	0.18	0.27	0.06	-0.13	0.19	0.28	0.12	0.08
Deviation [§]	0.37	0.42	0.50	0.38	0.40	0.50	0.41	0.41
Minimum Value	-1.00	-1.00	-0.88	-0.83	-0.59	-0.76	-0.57	-0.79
IR 0.05 Limit	-0.47	0.45	0.63	-0.79	-0.52	-0.80	-0.51	-0.67
IR 0.95 Limit	0.81	0.96	0.70	0.52	0.54	0.84	0.76	0.69
Maximum Value	1.00	1.00	-0.81	0.66	0.64	0.89	0.82	0.75
Kurtosis [¶]	-0.54	0.27	-1.29	-0.73	-1.40	-0.77	-1.09	-1.01
Skewness [¶]	0.21	-0.57	-0.22	0.04	-0.36	-0.49	-0.06	-0.15

Table S1 – Descriptive statistics for DS1 of both target and synthetic sample distributions when compared separately. Data in columns marked with * are reported using robust statistics; [‡]Metric for Gaussian distributions = mean, non-Gaussian = median; [§]Metric for Gaussian distributions = standard deviation, non-Gaussian = square root of the biweight midvariance; [¶]Variations include trimmed metrics for non-Gaussian distributions. Numbers marked in bold indicate the synthetic data that obtained the most significant rTOST equivalency p -values (Originally reported in Table 2).

	Original Data		LSGAN		WGAN		WGAN-GP	
	PC1*	PC2	PC1*	PC2	PC1	PC2*	PC1	PC2
Shapiro w	0.95	0.97	0.97	0.99	0.98	0.96	0.98	0.98
Shapiro p	0.02	0.15	0.006	0.39	0.09	0.002	0.171	0.08
Central Tendency [‡]	0.18	0.27	0.13	0.08	0.14	0.28	0.18	0.18
Deviation [§]	0.37	0.42	0.46	0.44	0.39	0.46	0.39	0.42
Minimum Value	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
IR 0.05 Limit	-0.47	0.45	-0.76	-0.70	-0.51	-0.66	-0.48	-0.53
IR 0.95 Limit	0.81	0.96	0.73	0.74	0.73	0.88	0.78	0.75
Maximum Value	1.00	1.00	1.00	1.00	1.00	1.00	-0.01	1.00
Kurtosis [¶]	-0.54	0.27	-0.44	-0.57	-0.05	-0.59	1.00	-0.45
Skewness [¶]	0.21	-0.57	-0.50	-0.13	-0.43	-0.42	-0.33	-0.34

Table S2 – Descriptive statistics for augmented DS1. Data in columns marked with * are reported using robust statistics; [‡]Metric for Gaussian distributions = mean, non-Gaussian = median; [§]Metric for Gaussian distributions = standard deviation, non-Gaussian = square root of the biweight midvariance; [¶]Variations include trimmed metrics for non-Gaussian distributions. Numbers marked in bold indicate the synthetic data that obtained the most significant rTOST equivalency p -values (Originally reported in Table 2).

	Original Data		LSGAN		WGAN		WGAN-GP	
	PC1*	PC2	PC1*	PC2	PC1*	PC2*	PC1*	PC2
Shapiro w	0.96	0.98	0.95	0.97	0.88	0.84	0.93	0.97
Shapiro p	0.05	0.31	0.03	0.11	1.9e-05	2.8e-06	0.002	0.15
Central Tendency [‡]	0.22	-0.16	0.15	-0.10	0.12	0.35	-0.07	-0.08
Deviation [§]	0.38	0.38	0.55	0.46	0.70	0.78	0.54	0.36
Minimum Value	-1.00	-1.00	-0.94	-0.90	-0.90	-0.97	-0.91	-0.68
IR 0.05 Limit	-0.69	-0.79	-0.80	-0.79	-0.88	0.91	-0.80	-0.59
IR 0.95 Limit	0.65	0.46	0.84	0.68	0.94	0.92	0.68	0.53
Maximum Value	1.00	1.00	0.92	0.85	0.94	0.94	0.83	0.84
Kurtosis [¶]	-0.27	0.59	-1.21	-1.01	-1.61	-1.66	-1.45	-0.65
Skewness [¶]	-0.59	0.52	-0.12	0.20	-0.09	-0.26	0.02	0.36

Table S3 – Descriptive statistics for DS2 of both target and synthetic sample distributions when compared separately. Data in columns marked with * are reported using robust statistics; [‡]Metric for Gaussian distributions = mean, non-Gaussian = median; [§]Metric for Gaussian distributions = standard deviation, non-Gaussian = square root of the biweight midvariance; [¶]Variations include trimmed metrics for non-Gaussian distributions. Numbers marked in bold indicate the synthetic data that obtained the most significant rTOST equivalency p -values (Originally reported in Table 2).

	Original Data		LSGAN		WGAN		WGAN-GP	
	PC1*	PC2	PC1*	PC2	PC1*	PC2*	PC1*	PC2
Shapiro w	0.96	0.98	0.97	0.98	0.95	0.95	0.96	0.98
Shapiro p	0.05	0.31	0.01	0.15	1.7e-04	1.5e-04	0.001	0.19
Central Tendency [‡]	0.22	-0.16	0.20	-0.13	0.20	-0.10	0.15	-0.12
Deviation [§]	0.38	0.38	0.48	0.42	0.57	0.61	0.50	0.37
Minimum Value	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
IR 0.05 Limit	-0.69	-0.79	-0.75	-0.79	-0.87	-0.91	-0.77	-0.67
IR 0.95 Limit	0.65	0.46	0.81	0.68	0.89	0.90	0.67	0.53
Maximum Value	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Kurtosis [¶]	-0.27	0.59	-0.81	-0.38	-0.95	-1.13	-1.05	0.03
Skewness [¶]	-0.59	0.52	-0.37	0.36	-0.36	0.22	-0.41	0.30

Table S4 – Descriptive statistics for augmented DS2. Data in columns marked with * are reported using robust statistics; [‡]Metric for Gaussian distributions = mean, non-Gaussian = median; [§]Metric for Gaussian distributions = standard deviation, non-Gaussian = square root of the biweight midvariance; [¶]Variations include trimmed metrics for non-Gaussian distributions. Numbers marked in bold indicate the synthetic data that obtained the most significant rTOST equivalency p -values (Originally reported in Table 2).

	Original Data		LSGAN		WGAN		WGAN-GP	
	PC1	PC2	PC1	PC2*	PC1	PC2*	PC1*	PC2*
Shapiro w	0.99	0.98	0.97	0.95	0.97	0.94	0.93	0.93
Shapiro p	0.75	0.30	0.13	0.02	0.15	0.004	0.002	0.002
Central Tendency [‡]	-0.10	-0.16	0.03	-0.29	-0.03	-0.002	-0.26	0.25
Deviation [§]	0.41	0.43	0.49	0.47	0.50	0.51	0.58	0.63
Minimum Value	-1.00	-1.00	-0.93	-0.91	-0.92	-0.86	-0.93	-0.95
IR 0.05 Limit	-0.81	-0.98	-0.78	-0.87	-0.77	-0.79	-0.90	-0.90
IR 0.95 Limit	0.47	0.61	0.78	0.49	0.82	0.76	0.76	0.93
Maximum Value	1.00	1.00	0.92	0.88	0.95	0.85	0.94	0.97
Kurtosis [¶]	-0.33	0.19	-1.00	-1.20	-1.01	-1.38	-1.23	-1.29
Skewness [¶]	-0.02	0.35	0.00	0.32	0.14	-0.03	0.33	-0.29

Table S5 – Descriptive statistics for DS3 of both target and synthetic sample distributions when compared separately. Data in columns marked with * are reported using robust statistics; [‡]Metric for Gaussian distributions = mean, non-Gaussian = median; [§]Metric for Gaussian distributions = standard deviation, non-Gaussian = square root of the biweight midvariance; [¶]Trimmed metrics for non-Gaussian distributions. Numbers marked in bold indicate the synthetic data that obtained the most significant rTOST equivalency p -values (Originally reported in Table 2).

	Original Data		LSGAN		WGAN		WGAN-GP	
	PC1	PC2	PC1	PC2	PC1	PC2*	PC1*	PC2*
Shapiro w	0.99	0.98	0.99	0.98	0.96	0.98	0.97	0.97
Shapiro p	0.75	0.30	0.36	0.09	0.25	0.04	0.01	0.01
Central Tendency [‡]	-0.10	-0.16	-0.04	-0.19	-0.07	-0.13	-0.16	-0.10
Deviation [§]	0.41	0.43	0.46	0.44	0.45	0.51	0.51	0.57
Minimum Value	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
IR 0.05 Limit	-0.81	-0.98	-0.79	-0.87	-0.79	-0.82	-0.89	-0.90
IR 0.95 Limit	0.47	0.61	0.74	0.58	0.67	0.76	0.72	0.86
Maximum Value	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Kurtosis [¶]	-0.33	0.19	-0.64	-0.38	-0.64	-0.90	-1.09	-0.96
Skewness [¶]	-0.02	0.35	0.07	0.38	0.13	0.22	0.04	0.16

Table S6 – Descriptive statistics for augmented DS3. Data in columns marked with * are reported using robust statistics; [‡]Metric for Gaussian distributions = mean, non-Gaussian = median; [§]Metric for Gaussian distributions = standard deviation, non-Gaussian = square root of the biweight midvariance; [¶]Variations include trimmed metrics for non-Gaussian distributions. Numbers marked in bold indicate the synthetic data that obtained the most significant rTOST equivalency p -values (Originally reported in Table 2).