A probabilistic method for generating α-shapes

Emily T. Winn-Nuñez

Joint Statistical Meetings, Toronto, ON

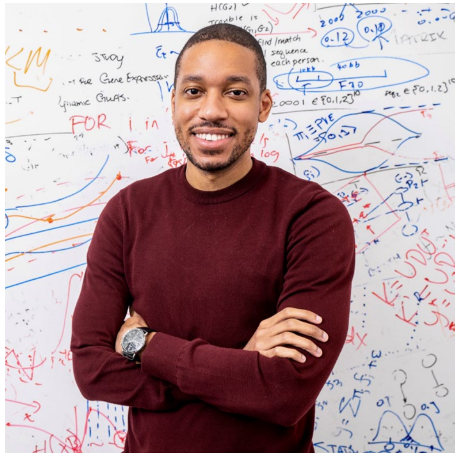
August 6, 2023

Website: etwinn.github.io
LinkedIn: emily-winn-math

Twitter: @EmilyTWinn13
BlueSky: @etwn13.bsky.social
Mastodon: @etwn@mathstodon.xyz

Collaborators

Dhananjay Bhaskar



Lorin Crawford

Ryan Huang

Jonathan Reichner

Hadley Witt

Ian Y. Wong

Outline

- Background, Intro to the math (shapes)
  - Why data augmentation important
  - Why alpha shapes
- Moving to a pipeline
- Results in 2D (Neutrophils)
- Results in 3D (Teeth)

Motivation

Shape is important to study in many fields….

- Biology (e.g., cancerous tumors, organ shape)
- Anthropology/morphology (e.g., femur shape, teeth)
- Computer graphics

… but shape data comes in small sets!

- Difficult to come by
- Expensive to collect and store

History of Shape Statistics

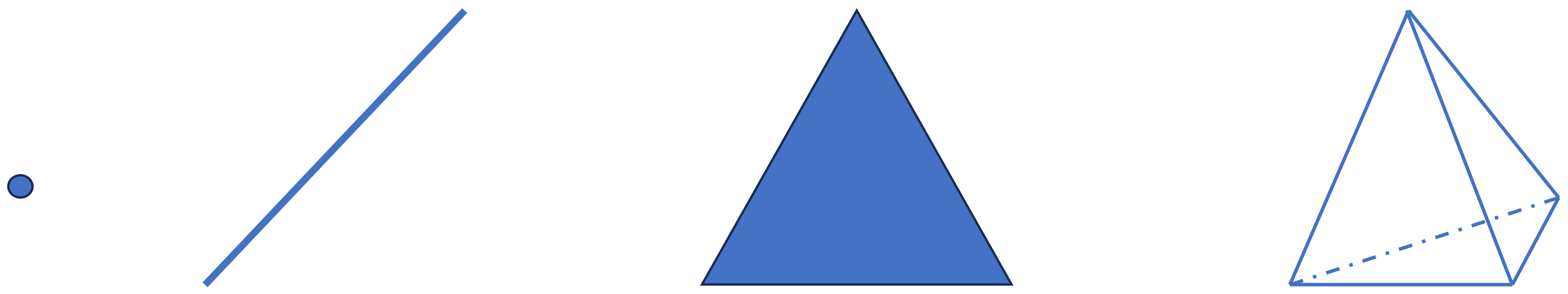
- When we build a new model, standard is to test it on simulated data and real data
- For shapes – there isn't really simulated data!
- Existing work in shape statistics
  - Statistics on landmarks (Albrecht et al. 2013), specific measurements, point clouds, summary statistics (Turner et al. 2014, Wang et al. 2022)
  - Sampling from manifolds/point clouds, shape reconstruction (Fasy et al. 2022)

(Albrecht et al., 2013)

What is a “Shape”?

(shape = simplicial complex representation of compact Riemannian Manifold embedded in Euclidean space)

In other words, anything I can approximate with topological “building blocks”



α-shapes

$Let\ B_\alpha(u)\ denote\ the\ ball\ of\ radius$
$\alpha\ at\ a\ point\ u \in S.$

$Let\ V(u)\ denote\ the\ Voronoi\ cell\ of\ u$

$Let\ R_u(\alpha) = B_\alpha(u) \cap V(u)$

$The\ union\ of\ R_u(\alpha) for\ all\ points\ u \in S$
$form\ a\ cover\ of\ S, the\ nerve\ of\ which$
$is\ the\ \boldsymbol{\alpha - complex}$

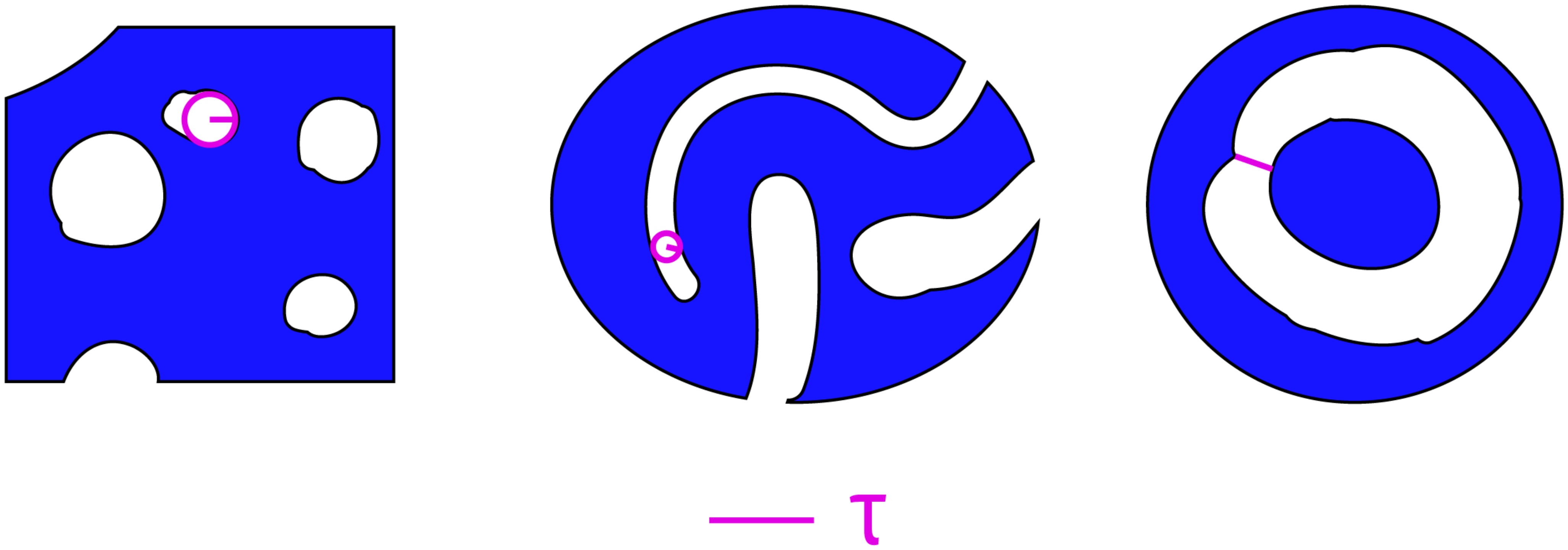
$The\ boundary\ of\ the\ \alpha - complex$
$defines\ the\ \boldsymbol{\alpha - shape}$

$\alpha = 0$

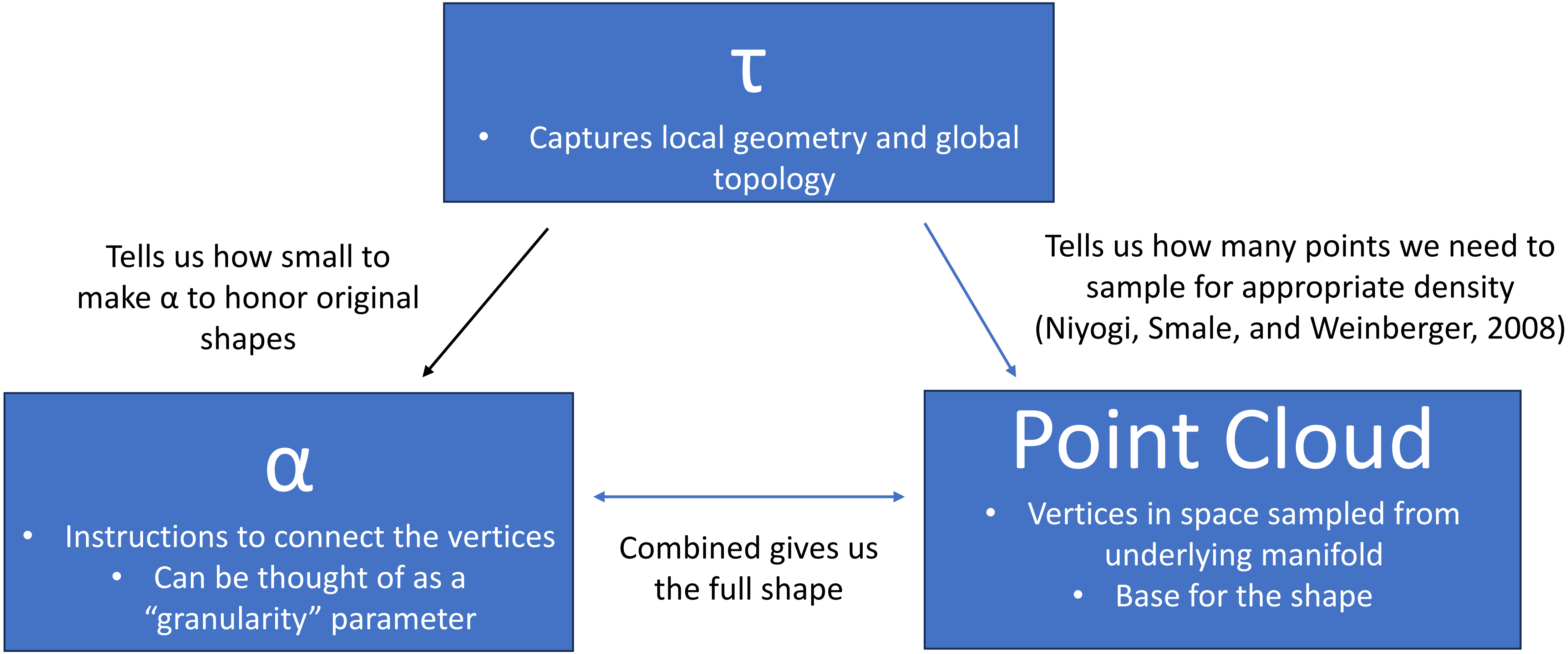
(Edulsbrunner and Harer)

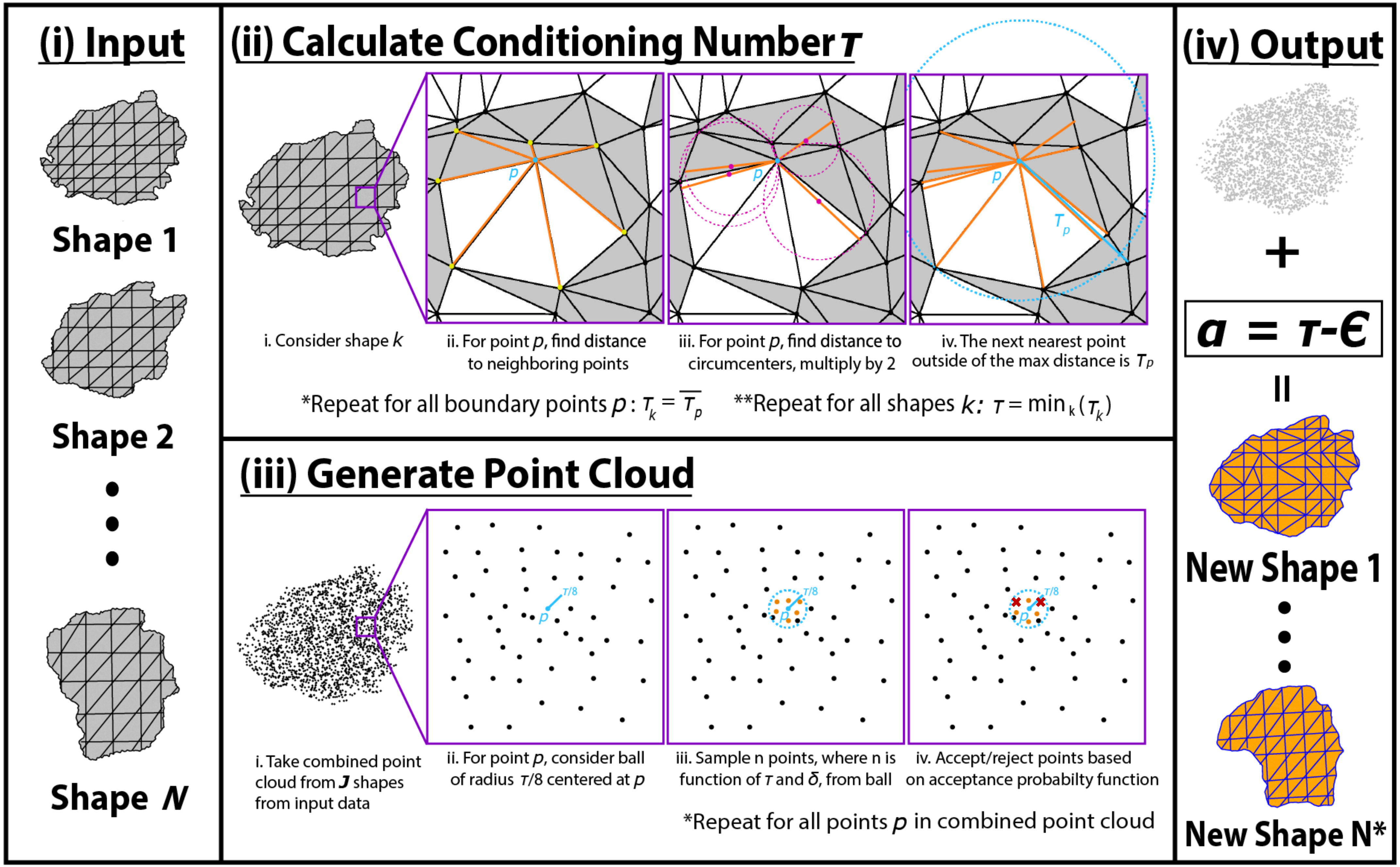
Conditioning Number τ

Represents the “tubulature” – a characteristic that captures the local geometry and global topology of a shape.

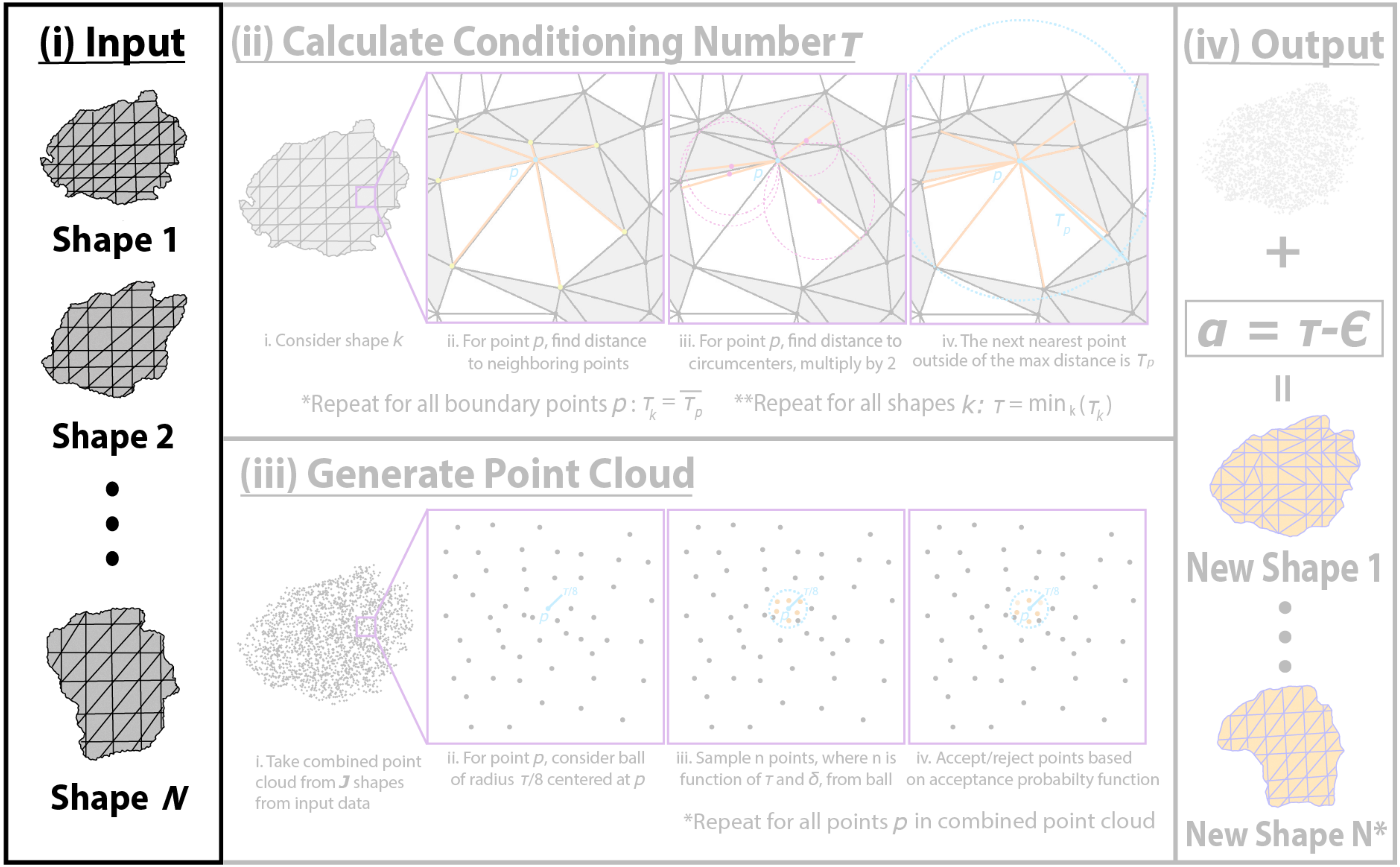


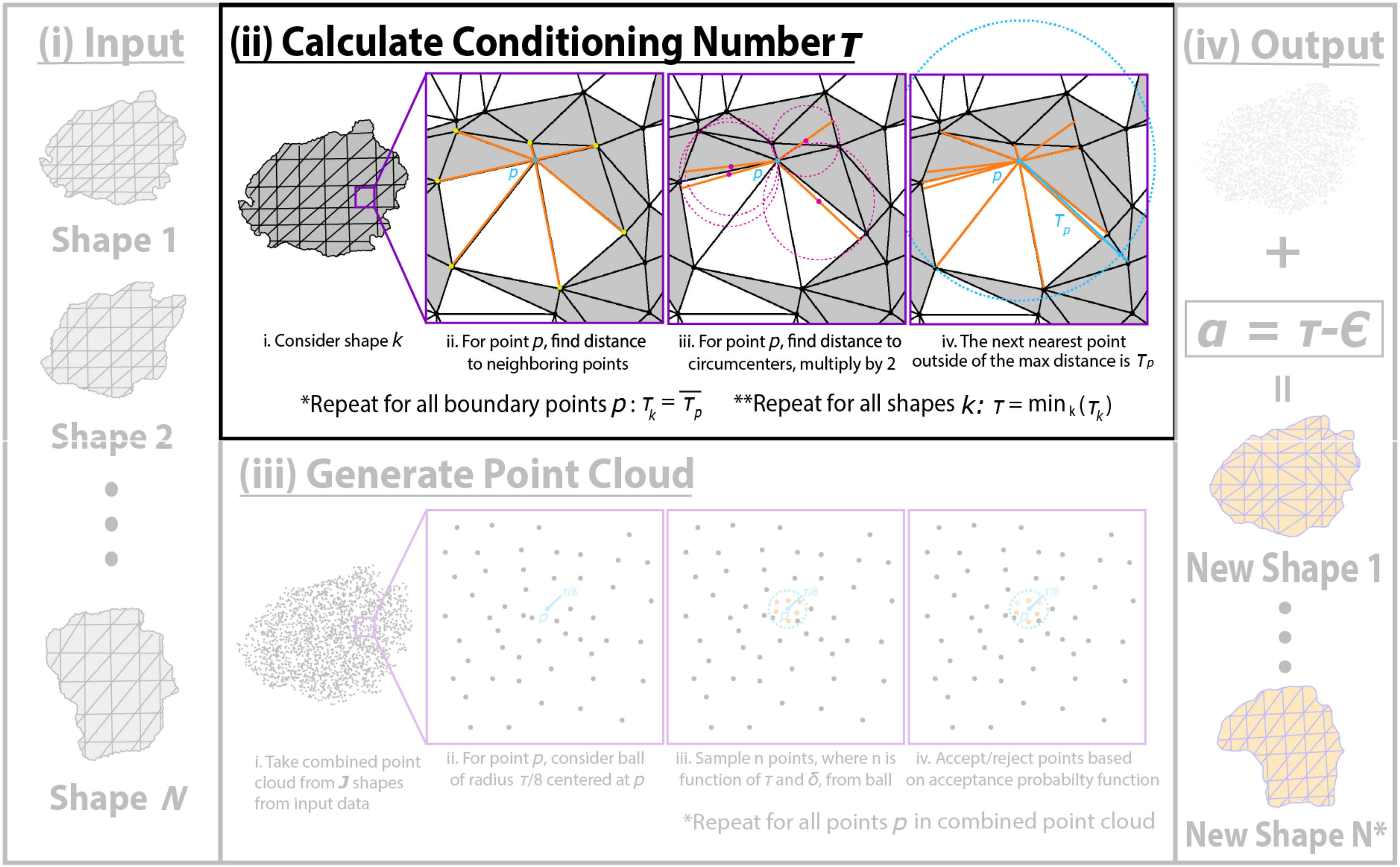
A probability model for alpha shapes

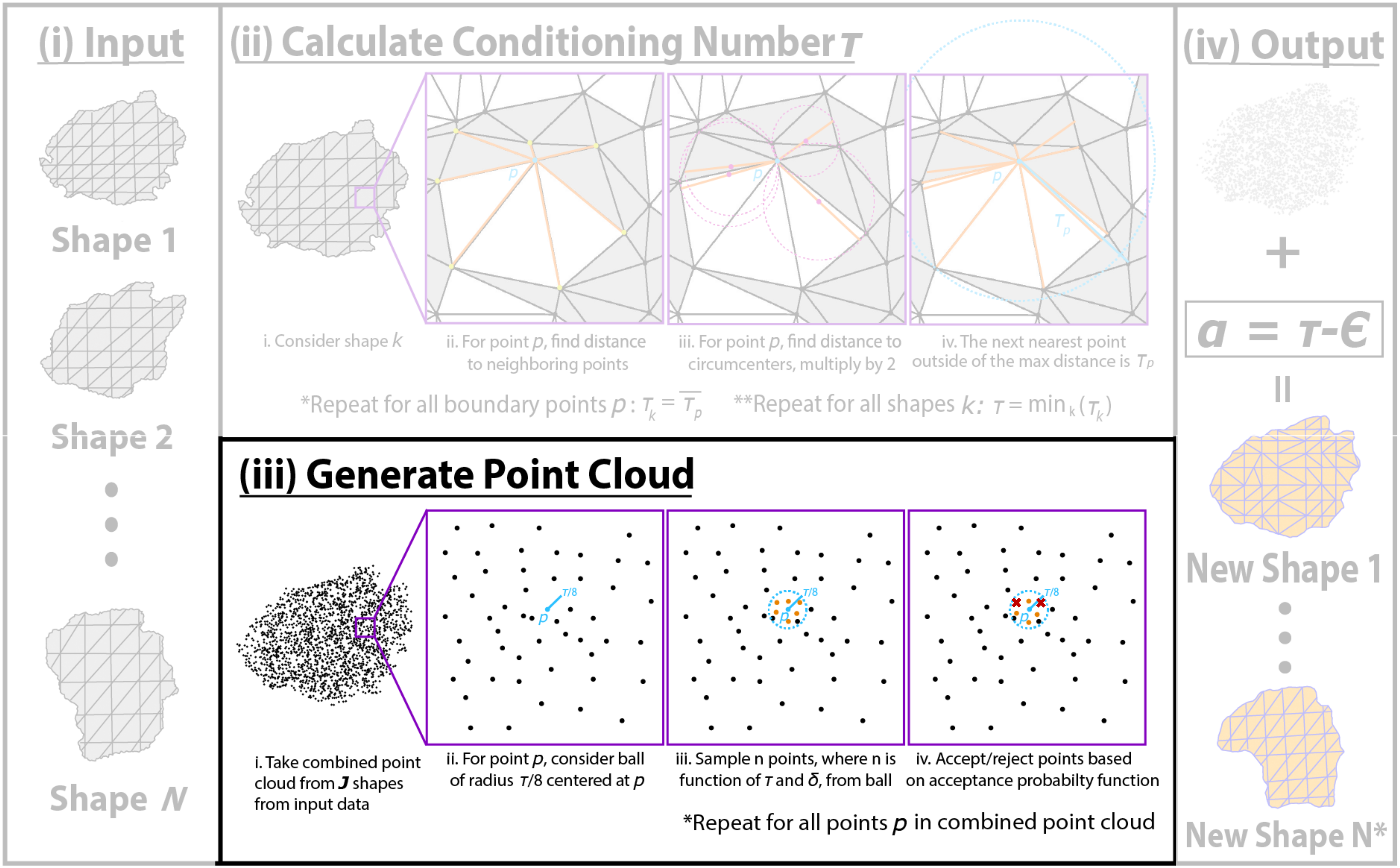




(i) Input
Shape 1
Shape 2
Shape N
(ii) Calculate Conditioning Number τ
p
i. Consider shape k
ii. For point p, find distance to neighboring points
iii. For point p, find distance to circumcenters, multiply by 2
iv. The next nearest point outside of the max distance is τp
*Repeat for all boundary points p : τk = mean(τp)
**Repeat for all shapes k: τ = mink(τk)
(iii) Generate Point Cloud
τ/8
i. Take combined point cloud from J shapes from input data
ii. For point p, consider ball of radius τ/8 centered at p
iii. Sample n points, where n is function of τ and δ, from ball
iv. Accept/reject points based on acceptance probabilty function
*Repeat for all points p in combined point cloud
(iv) Output
+
a = τ-ϵ
||
New Shape 1
New Shape N*


(i) Input
Shape 1
Shape 2
Shape N
(ii) Calculate Conditioning Number τ
i. Consider shape k
ii. For point p, find distance to neighboring points
iii. For point p, find distance to circumcenters, multiply by 2
iv. The next nearest point outside of the max distance is T_p
*Repeat for all boundary points p: $\tau_k = \overline{T_p}$
**Repeat for all shapes k: $\tau = \min_k(\tau_k)$
(iii) Generate Point Cloud
i. Take combined point cloud from J shapes from input data
ii. For point p, consider ball of radius τ/8 centered at p
iii. Sample n points, where n is function of τ and δ, from ball
iv. Accept/reject points based on acceptance probabilty function
*Repeat for all points p in combined point cloud
(iv) Output
$a = \tau - \epsilon$
New Shape 1
New Shape N*

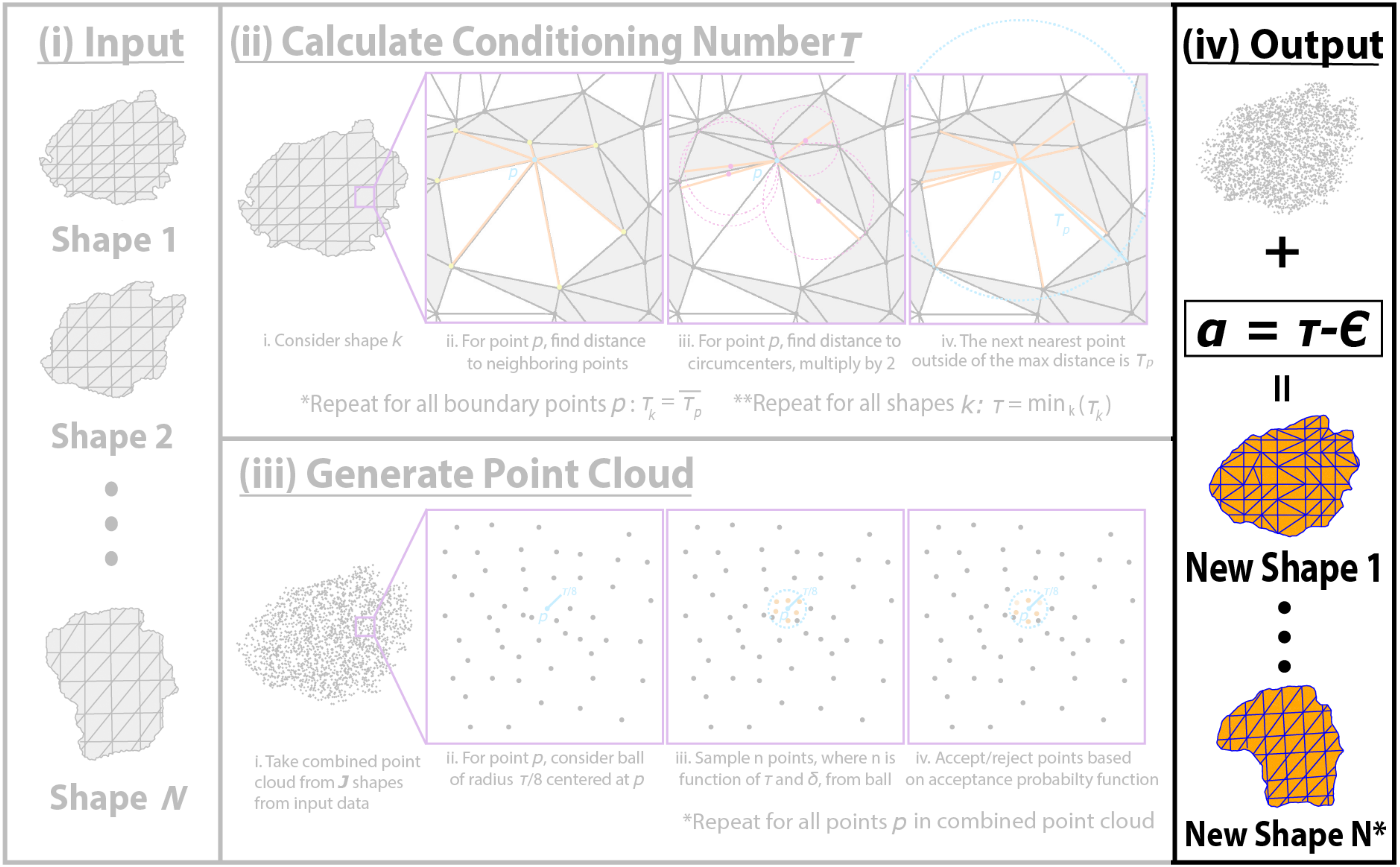

(i) Input
Shape 1
Shape 2
Shape N
(ii) Calculate Conditioning Number τ
p
T p
i. Consider shape k
ii. For point p, find distance to neighboring points
iii. For point p, find distance to circumcenters, multiply by 2
iv. The next nearest point outside of the max distance is Tp
*Repeat for all boundary points p : τk = T̄p
**Repeat for all shapes k: τ = mink(τk)
(iii) Generate Point Cloud
τ/8
p
i. Take combined point cloud from J shapes from input data
ii. For point p, consider ball of radius τ/8 centered at p
iii. Sample n points, where n is function of τ and δ, from ball
iv. Accept/reject points based on acceptance probabilty function
*Repeat for all points p in combined point cloud
(iv) Output
+
α = τ-ϵ
||
New Shape 1
New Shape N*


(i) Input
Shape 1
Shape 2
Shape N
(ii) Calculate Conditioning Number τ
i. Consider shape k
ii. For point p, find distance to neighboring points
iii. For point p, find distance to circumcenters, multiply by 2
iv. The next nearest point outside of the max distance is T_p
*Repeat for all boundary points p: $\tau_k = \overline{T_p}$
**Repeat for all shapes k: $\tau = \min_k(\tau_k)$
(iii) Generate Point Cloud
i. Take combined point cloud from J shapes from input data
ii. For point p, consider ball of radius τ/8 centered at p
iii. Sample n points, where n is function of τ and δ, from ball
iv. Accept/reject points based on acceptance probabilty function
*Repeat for all points p in combined point cloud
(iv) Output
α = τ-ϵ
New Shape 1
New Shape N*

Acceptance Probability for a Point *y*

$$P(\text{accept} \quad y) = \begin{cases} 0 & p(y) < \kappa \\ 1 - \exp(-\frac{2}{\kappa * J}(p(y) - \kappa)) & \kappa \leq p(y) < \kappa * J \\ 1 & p(y) \geq \kappa * J \end{cases}$$

- *y* is the proposed new point to add to the generated point cloud
- p(*y*) is number of points in base point cloud within radius **τ**/4 of *y*
- ***J*** is number of shapes from original set selected for base point cloud.
- **κ** is the minimum number of points from the base point cloud that we require to be within radius **τ**/4 of new point *y*

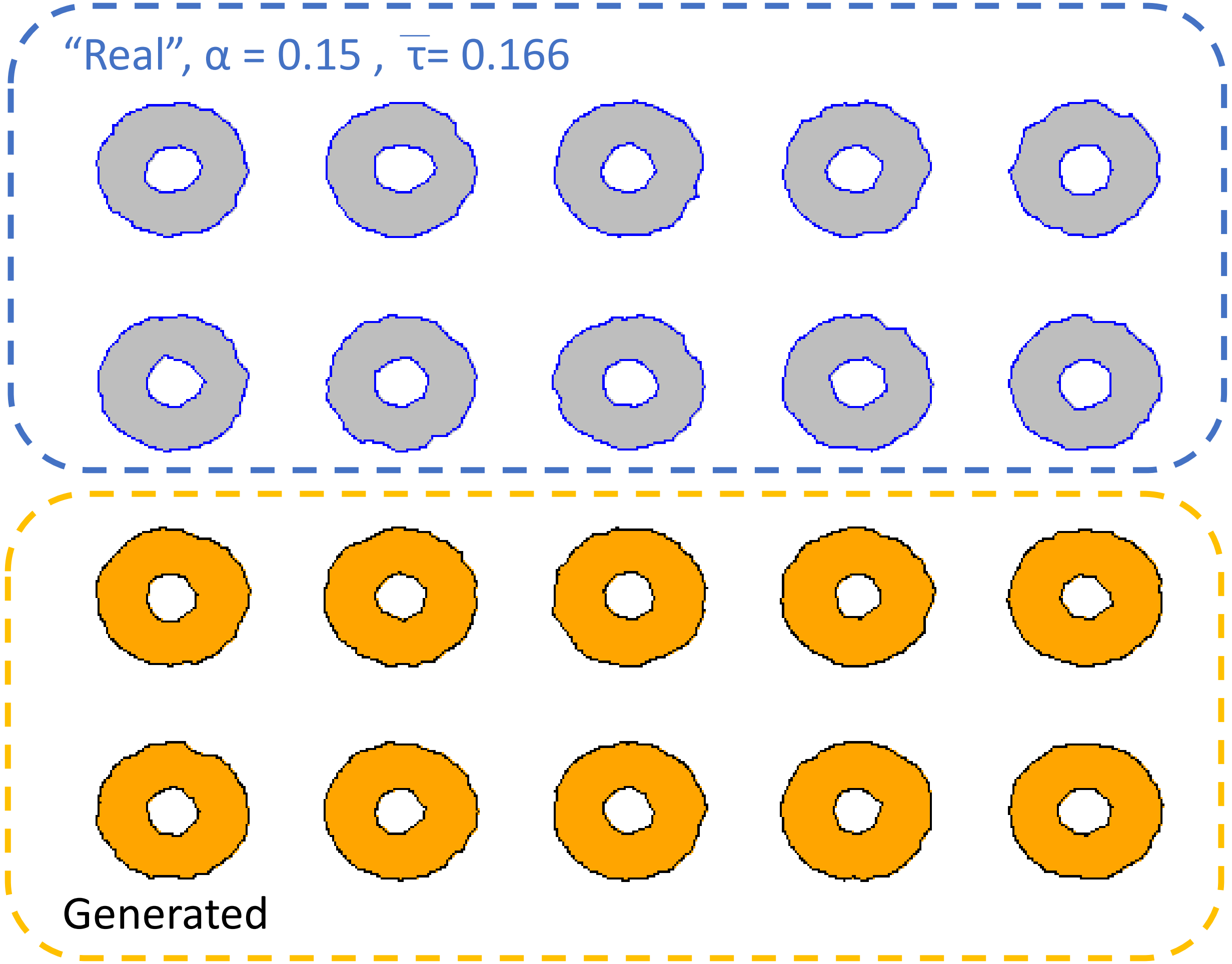


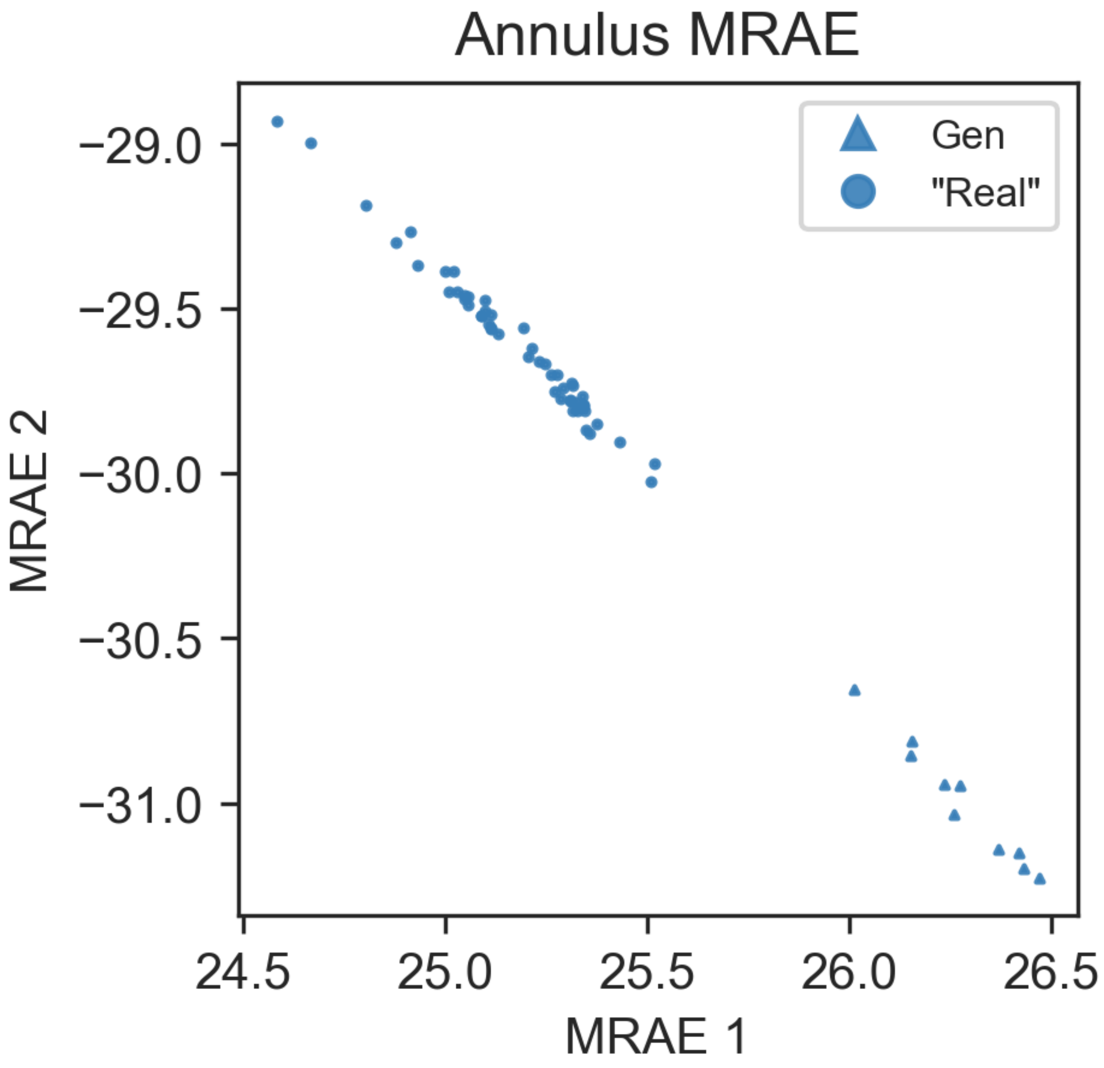
(i) Input
Shape 1
Shape 2
Shape N
(ii) Calculate Conditioning Number τ
i. Consider shape k
ii. For point p, find distance to neighboring points
iii. For point p, find distance to circumcenters, multiply by 2
iv. The next nearest point outside of the max distance is Tp
*Repeat for all boundary points p : τk = Tp
**Repeat for all shapes k: τ = mink(τk)
(iii) Generate Point Cloud
i. Take combined point cloud from J shapes from input data
ii. For point p, consider ball of radius τ/8 centered at p
iii. Sample n points, where n is function of τ and δ, from ball
iv. Accept/reject points based on acceptance probabilty function
*Repeat for all points p in combined point cloud
(iv) Output
+
a = τ-ϵ
New Shape 1
New Shape N*

Toy Example: Annuli

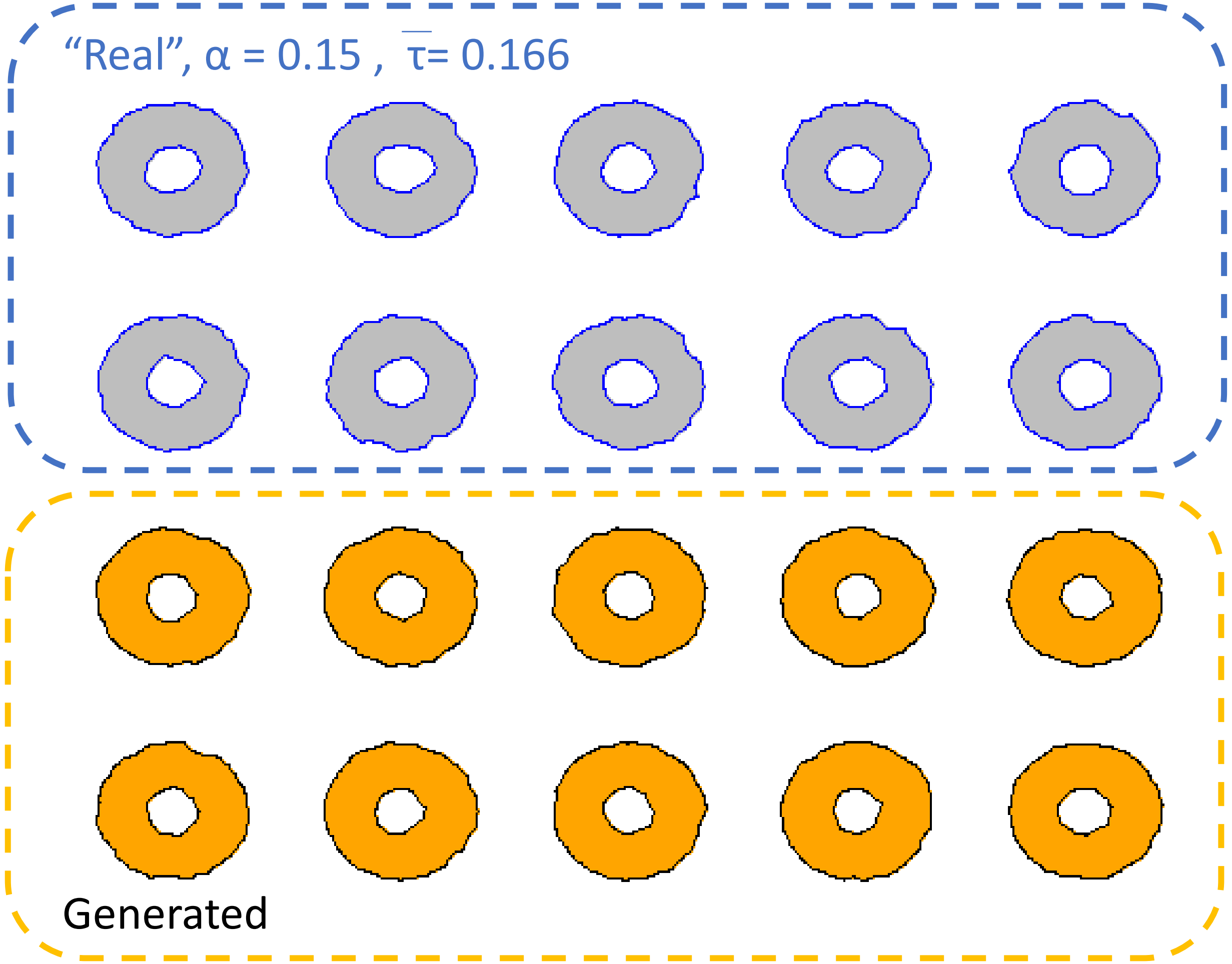
- The Data:
  - 50 simulated annuli, as α-shapes
  - Each annulus is 500 points, sampled between radius 0.25 and 0.75, $\alpha = 0.15$
- The Analysis:
  - 43 characteristics for 2D shapes measured – e.g., area, perimeter, centroids
  - Manifold Regularized AutoEncoder (MRAE) used for dimension reduction of shape characteristic vectors.

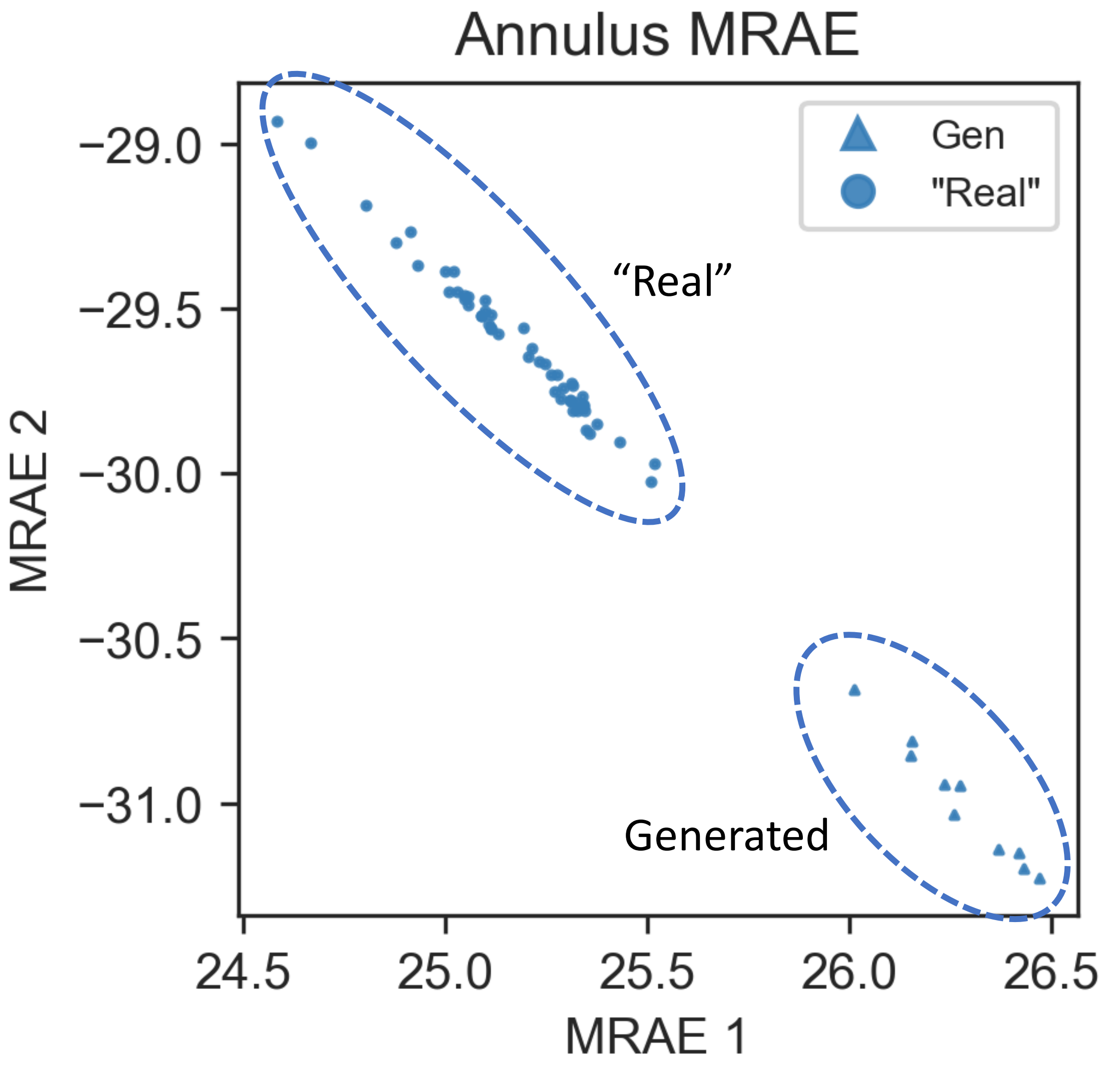
Annuli Results





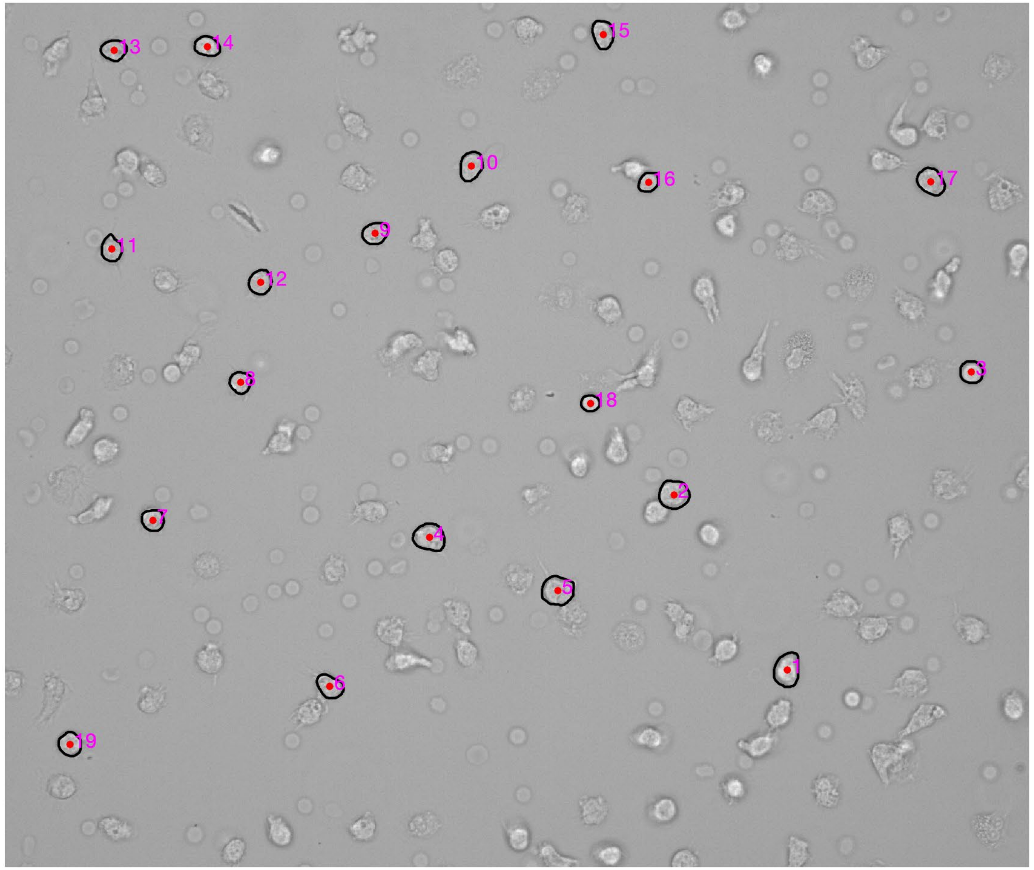
Annuli Results





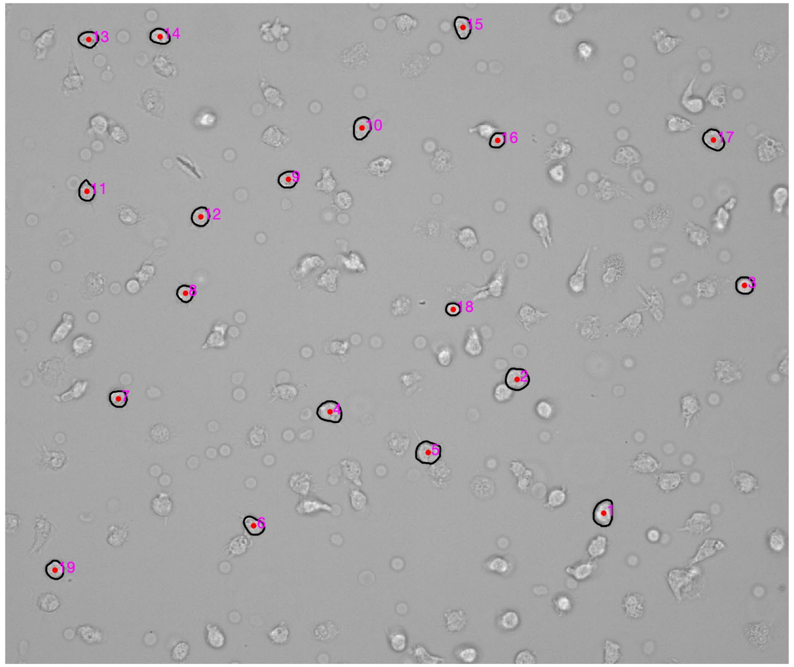
2D Shapes: Healthy and Septic Neutrophils

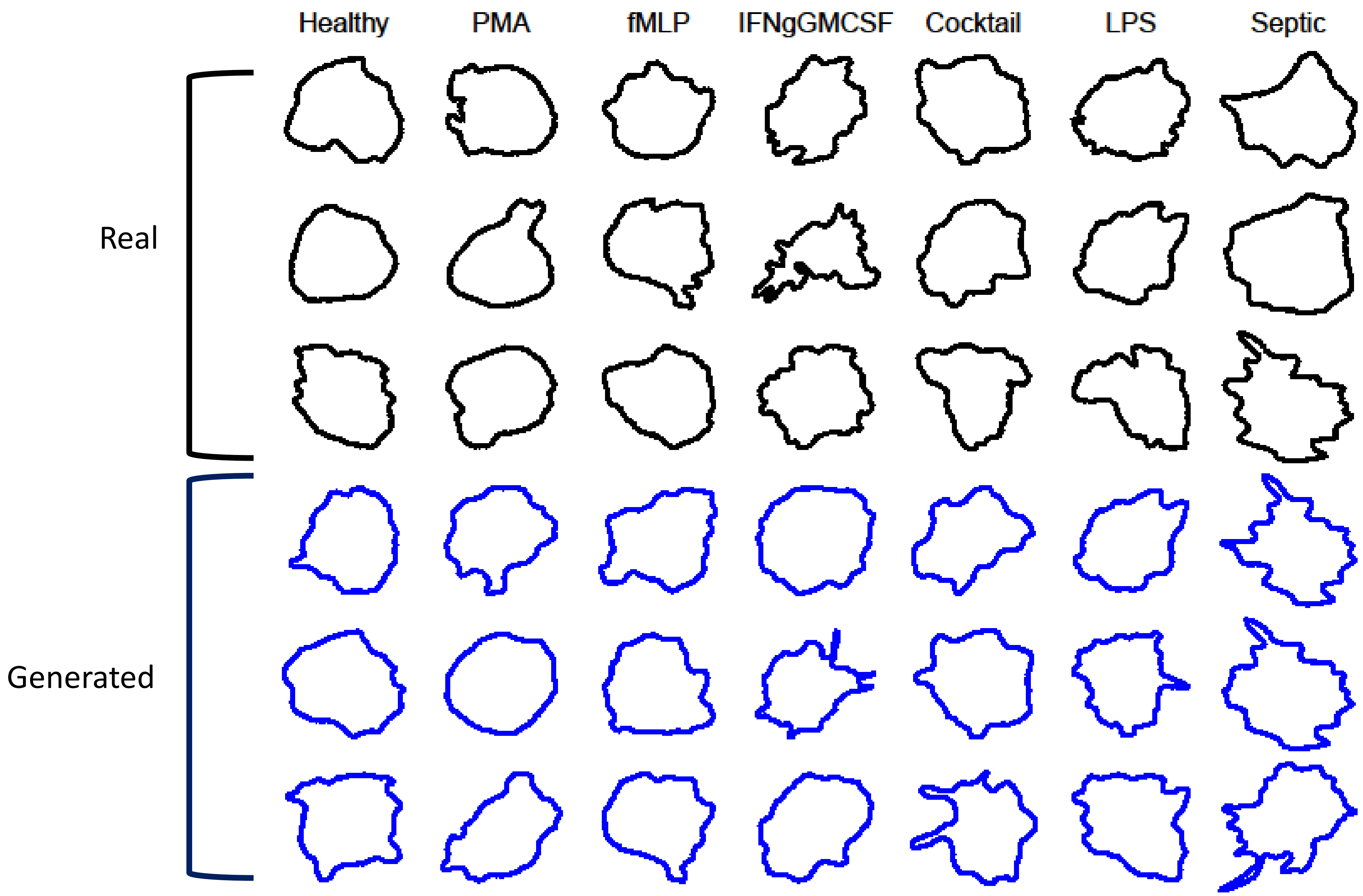
- Neutrophils are a type of white blood cell
- Immunologists study the change in shape with addition of a stimulant to shed insight on immune system response
- Of particular interest are septic neutrophils – which are incredibly hard to collect



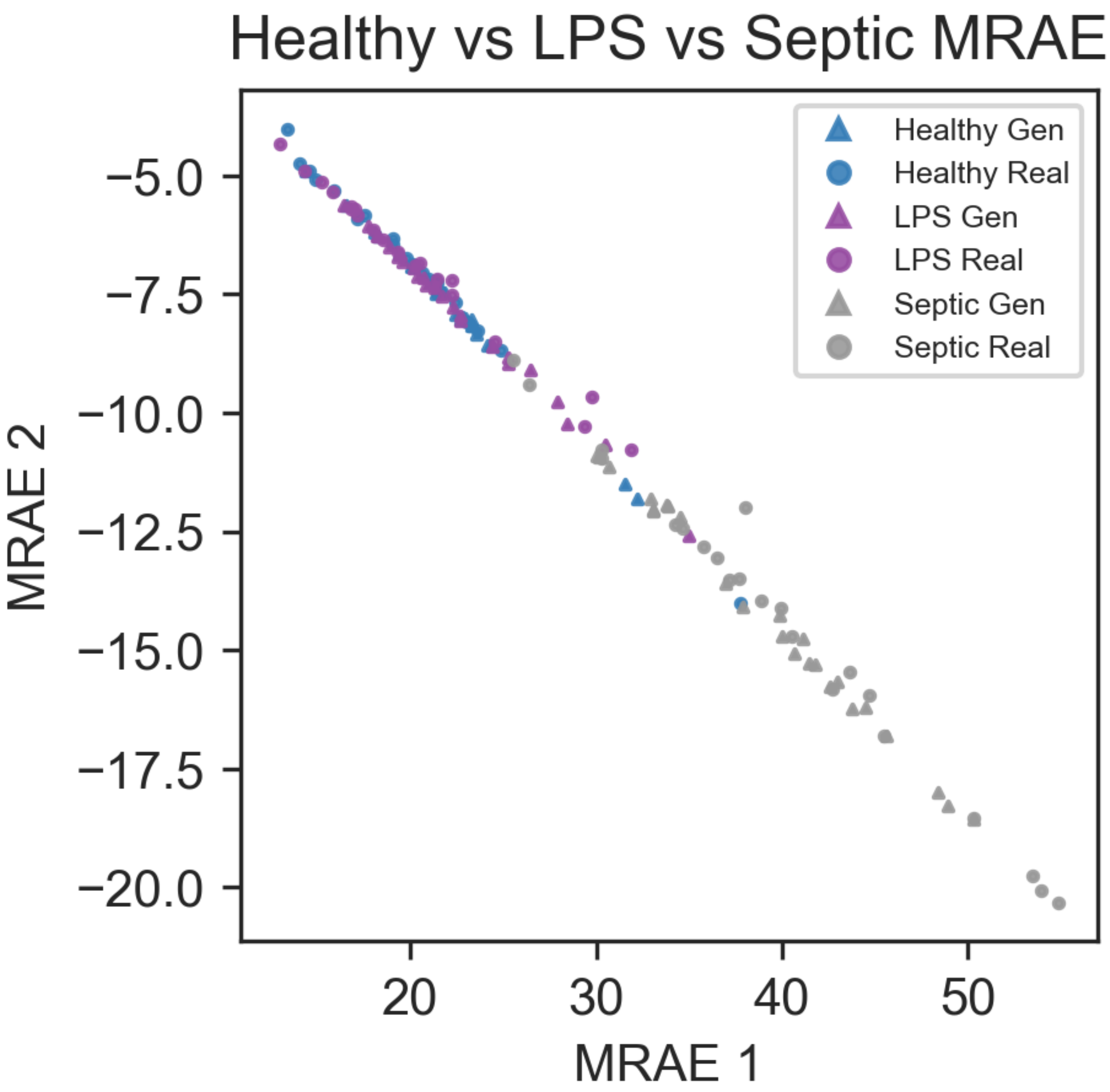
2D Shapes: Healthy and Septic Neutrophils

- The Data:
  - Neutrophils from healthy human tissues with seven stimulants added
  - Shape recorded before adding stimulant and 30 minutes after adding stimulant
  - Given as binary masks, converted to simplicial complexes for the pipeline
- The Analysis:
  - 43 characteristics for 2D shapes measured – e.g., area, perimeter, centroids
  - Manifold Regularized AutoEncoder (MRAE) used for dimension reduction of shape characteristic vectors.



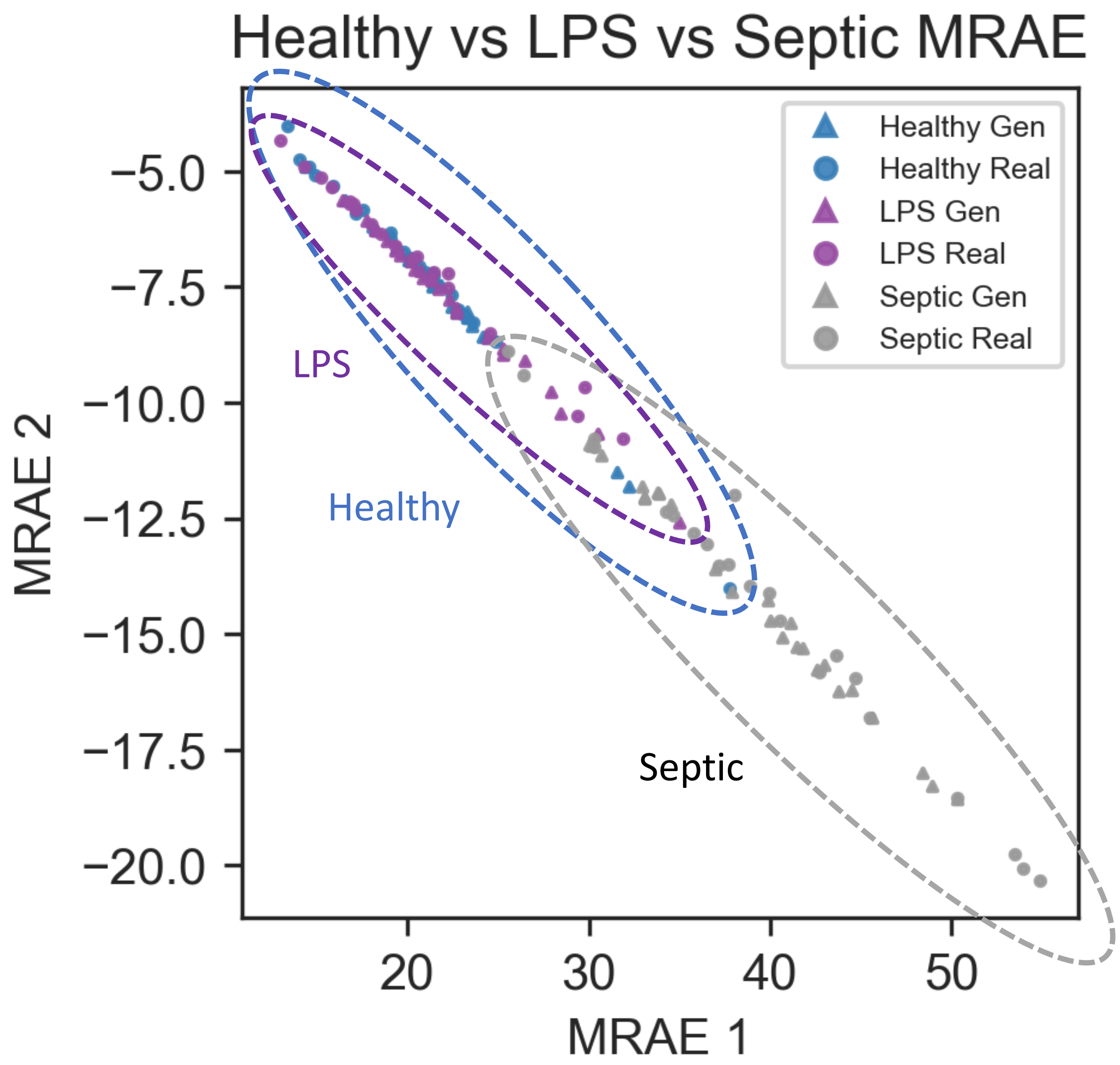

Healthy
PMA
fMLP
IFNgGMCSF
Cocktail
LPS
Septic
Real
Generated

Results



Healthy vs LPS vs Septic MRAE
Healthy Gen
Healthy Real
LPS Gen
LPS Real
Septic Gen
Septic Real
−5.0
−7.5
−10.0
−12.5
−15.0
−17.5
−20.0
MRAE 2
20
30
40
50
MRAE 1

Results

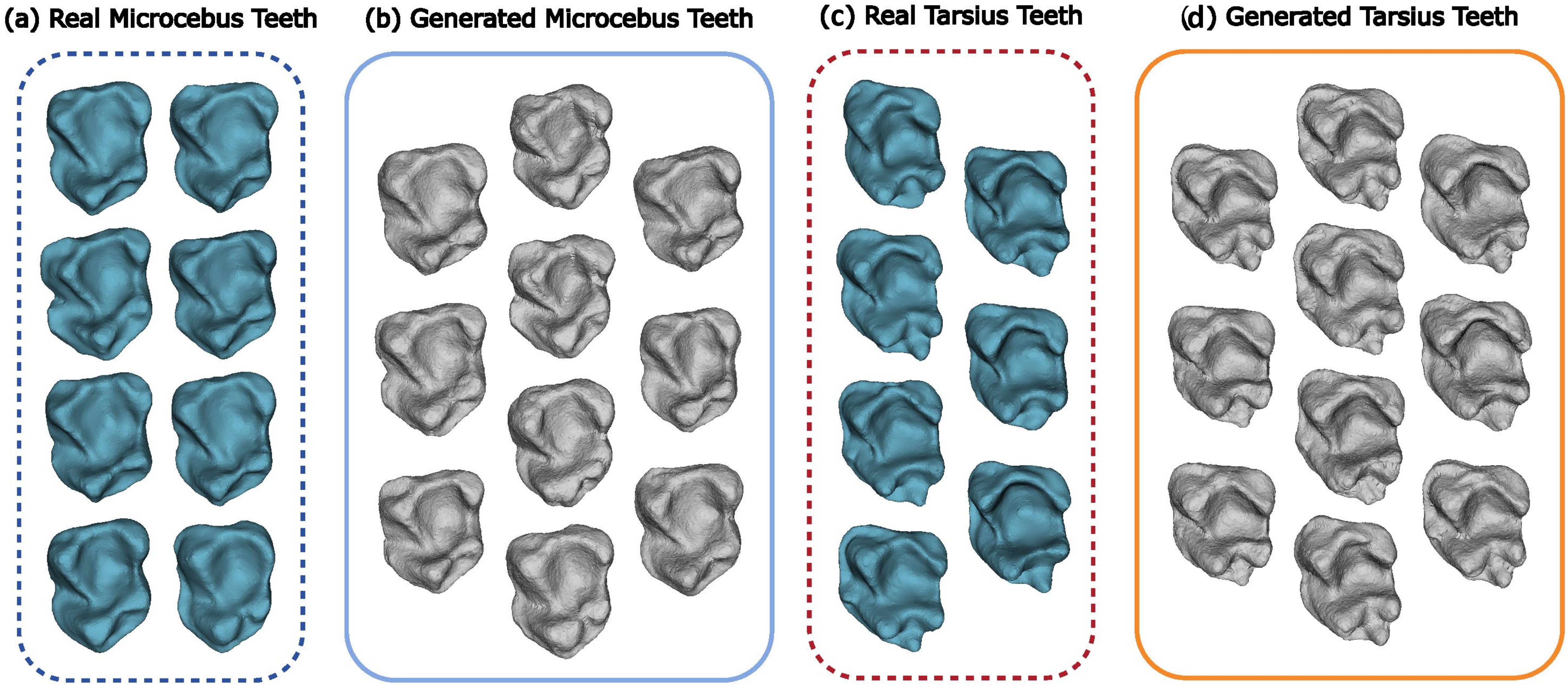

Healthy vs LPS vs Septic MRAE
Healthy Gen
Healthy Real
LPS Gen
LPS Real
Septic Gen
Septic Real
LPS
Healthy
Septic
MRAE 2
−5.0
−7.5
−10.0
−12.5
−15.0
−17.5
−20.0
20
30
40
50
MRAE 1

3D Shapes: Primate Teeth

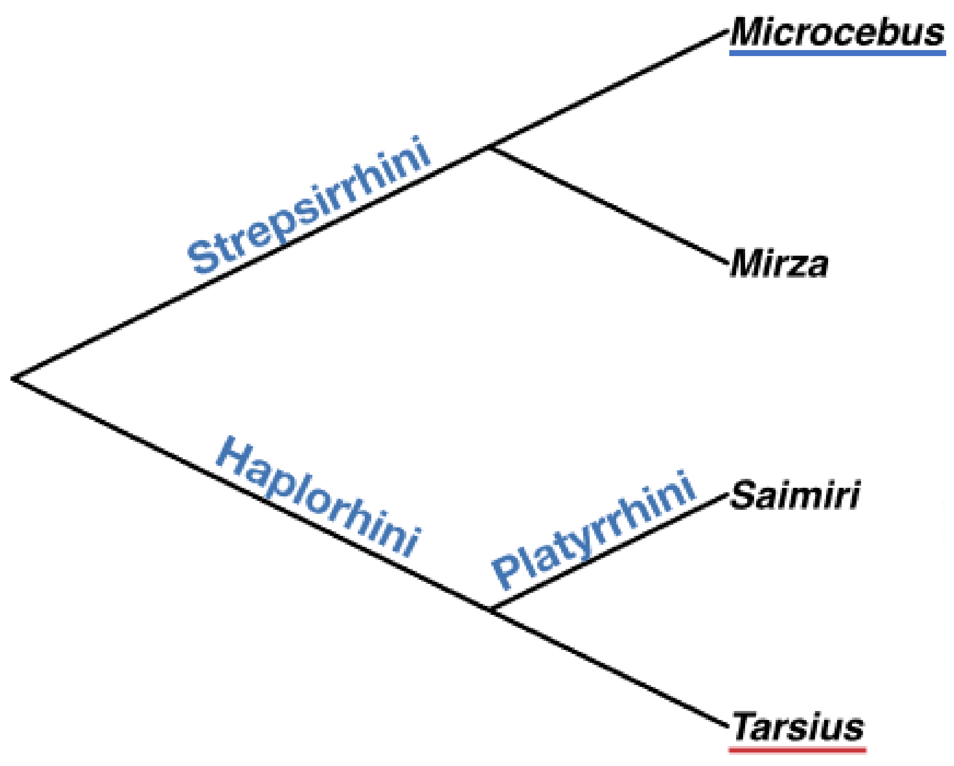
- The Data:
  - CT scans of mandibular molars from two different primate species, *Microcebus* and *Tarsius*
  - Teeth prealigned and scaled before sending through pipeline

- The Analysis:
  - Procrustes Analysis (Gower, 1975) via `auto3dgm` (Puente, 2013) assigns 400 landmark points and aligns based on size and scale
  - Uniform Manifold Approximation Projection (UMAP) (McIness et al., 2018) shows how the data cluster according to landmarks.

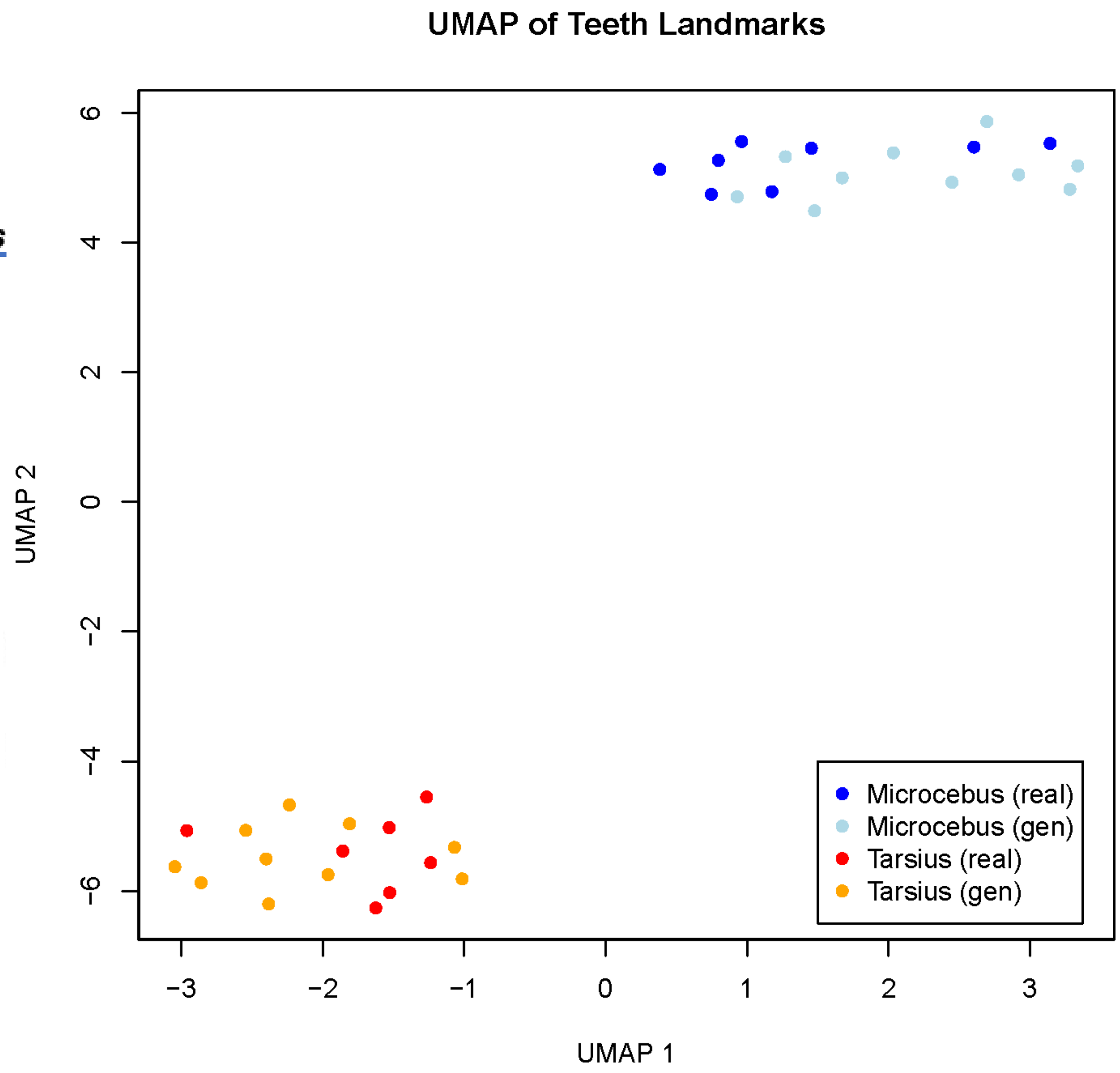
Results


(a) Real Microcebus Teeth
(b) Generated Microcebus Teeth
(c) Real Tarsius Teeth
(d) Generated Tarsius Teeth

Results



Phylogenetic Tree (Wang et al., 2021)



Conclusion and Further Work

- Presented a pipeline for generating new shapes to augment existing data sets
- Work to be extended to weighted alpha shapes, which have different applications (e.g., DNA, protein, where vertices are interpretable)
- `ashapesampler` R package will be available on Github.

Acknowledgements

Crawford Lab.

- Lorin Crawford (PI)
- Alan DenAdel
- Ryan Huang
- Chib Nwizu
- Julian Stamp
- Ria Vanod
- Helen Xie
- Alexandra (Alex) Wong

Funding support:

- National Science Foundation Graduate Research Fellowship Program, Grant No. 1655760
- Division of Applied Mathematics, Brown University

NSF BROWN

References

Albrecht T, Lüthi M, Gerig T, Vetter T. Posterior shape models. Medical Image Analysis;17(8):959-973. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1361841513000844

Edelsbrunner H, Harer J. Computational topology: an introduction. American Mathematical Society;. OCLC: ocn427757156.

Fasy BT, Komendarczyk R, Majhi S, Wenk C. On the Reconstruction of Geodesic Subspaces of N. International Journal of Computational Geometry & Applications. 2022;32(1):91–117. Available from: https://www.worldscientific.com/doi/10.1142/S0218195922500066.

Gower JC. Generalized procrustes analysis. Psychometrika;40(1):33–51. Available from: http://link.springer.com/10.1007/BF02291478.

McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. Journal of Open Source Software;3(29):861. Available from: http://joss.theoj.org/papers/10.21105/joss.00861.

Puente, J. Distances and algorithms to compare sets of shapes for automated biological morphometrics. Princeton University; 2013. Available from: http://arks.princeton.edu/ark:/88435/dsp01sq87bt73n.

Thompson DW. On growth and form. Cambridge University Press; 1917.

Turner K, Mukherjee S, Boyer DM. Persistent homology transform for modeling shapes and surfaces. Information and Inference;3(4):310–344. Available from: https://academic.oup.com/imaiai/article-lookup/doi/10.1093/imaiai/iau011.

Wang B, Sudijono T, Kirveslahti H, Gao T, Boyer DM, Mukherjee S, et al. A statistical pipeline for identifying physical features that differentiate classes of 3D shapes. The Annals of Applied Statistics. 2021;15