



DEPARTMENT OF COMPUTER SCIENCE

3D Individual Corallite Reconstruction
via Volumetric Vision Transformers and Topological Loss

Rob Jones

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree
of Bachelor of Science in the Faculty of Engineering.

Wednesday 3rd May, 2023

Abstract

Coral reefs are among the most densely populated areas of biodiversity in our oceans. They serve as a critical component of the oceanic ecosystem by providing nutrition and a safe haven for a wide range of species. However, climate change poses a significant threat to their survival. Warming and acidification of seas and oceans have increased the risk of coral extinction within this century. To ensure their continued existence, it is essential to understand the history of corals in their environment.

The work in this thesis presents, for the first time, an end-to-end automated pipeline for the three-dimensional reconstruction of individual corallites. Leveraging a fine-tuned, state-of-the-art volumetric segmentation model, and introducing a new, novel, topological loss function we show corallite segmentation quality between 0.63 and 0.77 DSC for completely unseen data, across multiple species of the *Porites* genus. Using cross-sectional slices of colony skeletons of the *Porites* genus, constructed from micro-computed tomography scans, volumetric inputs are generated and used train a transformer-based deep learning model to produce a segmented volume of all corallite regions of a given colony. The output segmentations are used as input to a automated modelling system used to generate a 3D representation of each unique corallite. This novel work is an important first step towards a comprehensive spatial-temporal corallite reconstruction system suitable for interactive visualisation of historic growth.

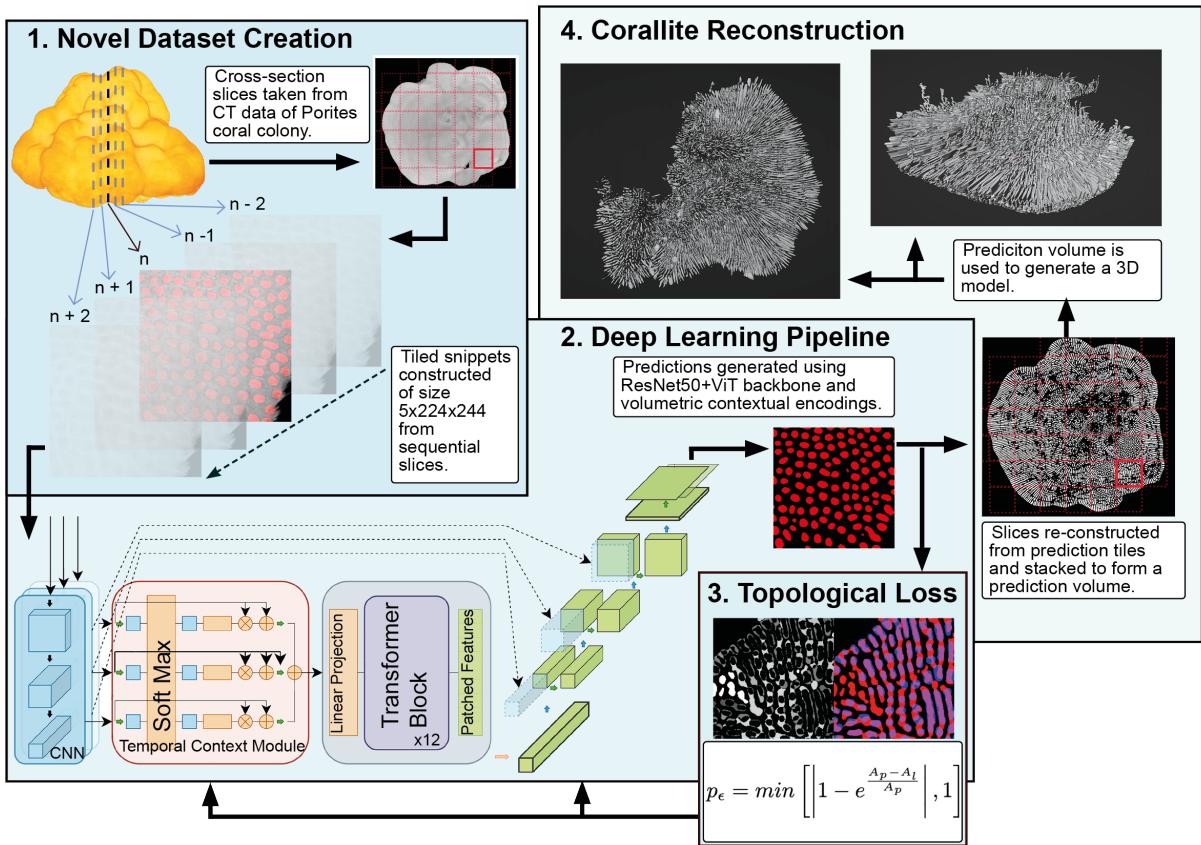


Figure 1: Project Overview: (1) The data is prepared as tiled, volumetric 'snippets' from sequential μ CT slice data. (2) The UNet-TCM-ViT deep learning pipeline is used to generate a segmentation prediction for each input tile. (3) Custom topological loss encourages better corallite region segmentation. Left: Loss map where pixel value is proportionate to error. Right: Composite ground-truth (blue) and prediction (red). (4) Predictions are reformed into full slices. These are stacked to form a prediction volume, serving as input to generate a fully interactive 3D model. Deep-learning pipeline diagram used in (2) adapted from [42]. Raw data sources for all stages provided courtesy of University of Bristol Earth Sciences as part of the ongoing 4DReef Project: <https://www.4d-reef.eu/>.

Dedication and Acknowledgements

I would like to thank my supervisor Dr. Tilo Burghardt for his invaluable insights, continuous motivation and enjoyable conversation throughout the duration of this project.

To my family, and friends, who helped with some of the annotation work and generally encouraged me to keep challenging myself and improve on this work.

And most importantly, to Úna, who's made huge changes to her life, put up with me, and supported me over the last three years of this degree (and beyond). I couldn't have done it without you.

Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Taught Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, this work is my own work. Work done in collaboration with, or with the assistance of others, is indicated as such. I have identified all material in this dissertation which is not my own work through appropriate referencing and acknowledgement. Where I have quoted or otherwise incorporated material which is the work of others, I have included the source in the references. Any views expressed in the dissertation, other than referenced material, are those of the author.

RA Jones

Rob Jones, Wednesday 3rd May, 2023

Contents

1	Introduction	1
1.1	Goals and Objectives	1
1.2	Contributions	2
2	Background	3
2.1	Coral Ecology	3
2.2	Technical Background	5
2.3	Previous Work	12
3	Project Implementation and Execution	14
3.1	Novel Volumetric Data	14
3.2	Deep Learning Pipeline	21
3.3	Corallite Reconstruction	30
4	Results and Critical Evaluation	37
4.1	Novel Volumetric Data	37
4.2	Deep-Learning Pipeline	39
4.3	Corallite Reconstruction	49
4.4	Results and Evaluation Summary	53
5	Conclusion	54
5.1	Project Status	54
5.2	Future Opportunities	56

List of Figures

1	Project Overview: (1) The data is prepared as tiled, volumetric 'snippets' from sequential μ CT slice data. (2) The UNet-TCM-ViT deep learning pipeline is used to generate a segmentation prediction for each input tile. (3) Custom topological loss encourages better corallite region segmentation. Left: Loss map where pixel value is proportionate to error. Right: Composite ground-truth (blue) and prediction (red). (4) Predictions are reformed into full slices. These are stacked to form a prediction volume, serving as input to generate a fully interactive 3D model. Deep-learning pipeline diagram used in (2) adapted from [42]. Raw data sources for all stages provided courtesy of University of Bristol Earth Sciences as part of the ongoing 4DReef Project: https://www.4d-reef.eu/	i
2.1	Corallites: A selection of images presenting corallites from a variety of perspectives. Left: A section of a Porites Compressa colony. Center Close-up of the same section illustrating the distribution and density of the corallites. Right: Diagram of the structure of a single corallite. Note the size of the corallite region is 1.4 - 1.8mm only. Image credits: Corals of the World.	4
2.2	μCT Slices: An illustrative representation of how a cross-sectional slice is taken from μ CT data of the full colony skeleton. The projection axis, throughout the context of the discussion in this thesis, is perpendicular to the plane formed by the slice. Image credit [32].	5
2.3	Artificial Neural Network: A diagram of a fully connected artificial neural network. Inputs $x_i \in X$ are fed through two hidden layers consisting activation functions based on a weighted sum of the inputs. The activation is triggered if and only if that weighted sum passes some threshold. The output layers represent probabilities that the input belong to each of the N classes, in this case $N = 2$. Image credit: DevSkrol.	6
2.4	Convolution Kernels: A toy convolution kernel operation in which the output is the sum of the element-wise multiplication of the kernel and its window. Image credit: Towards Data Science.	7
2.5	Max Pooling: An toy example of a max pooling operation with patch size 2×2 . Note the reduction of each patch is simply the maximum value within that section.	7
2.6	Bi-Linear Sampling: This image depicts bi-linear down-sampling from Source to Intermediate, followed by bi-linear up-sampling from Intermediate to Target. The lines represent contributions to the given interpolated element in the direction of the down-facing arrow. Image credit: Bart Wronski.	8
2.7	UNet Architecture: An example UNet architecture. Encoding occurs on the left side of the architecture, where the input is down-sampled. The bottle-neck is the central section. Decoding occurs on the right side when the outputs of the bottle neck are up-sampled to the original input size. The grey horizontal arrows represent <i>skip connections</i> , discussed in section 2.2.6. Image credit Towards Data Science.	8
2.8	Topological Loss Visualisation: An example topological error map (left) where brightness indicates larger error. Corresponding prediction/ground-truth overlay (right) where predictions are red, ground truth blue and therefore, correct pixel-wise predictions present as purple. Note that the errors are of a topological nature, a result of false positive predictions forming connections between corallite regions.	9
2.9	Transformer Component Architecture: (1) Transformer encoder/decoder components. (2) Multi-Head Attention (MHA) component. (3) Scaled Dot-Product Attention (SDPA) component, as described in [37]. The MHA provides contextual encoding through concatenation and linear projection of a number of sequential SDPA calculations, where Q, K and V represent Query, Key and Value respectively.	10
2.10	Vision Transformer Architecture: An illustrative example of the ViT architecture used in [10]. Note the input is separated into a sequence of patches, and associated positional embeddings. These patches are flattened by linear projection before being fed into the traditional transformer network architecture.	11
2.11	Skip Connections and Loss-landscape Visualisations: The impacts of skip-connections on the loss landscapes of ResNet56, visualised by [22]. Note the how the resulting landscape is much smoother, simplifying the process of finding a global minimum.	12

2.12 Video Trans-UNet Architecture: The architecture diagram of the Video Trans-UNet pipeline, as presented in [42].	13
3.1 Corallite Region Variations: An illustrative example of the differences in growth direction within a μ CT slice. Circular structures (red) represent growth aligned to the projection axis, elliptical structures (blue) represent growth perpendicular to the projection axis	16
3.2 Inherited Annotation Set: An example of the annotations inherited by this project. Top-Left: Raw μ CT data, Bottom Left: Annotations, Right: Composite where annotations are presented in red. Note that the annotations only cover a subset of the total number of corallite regions, specifically, only those with circular geometric shape.	16
3.3 Novel Annotation Set: Additional annotations produced for this project. Left: μ CT scan, Right: Novel Annotation. Top to Bottom: Porites Astraeiformis, Porites Naturalis 6781 Growth slice#1279, Porites Naturalis 6781 Growth slice#2499, Porites Naturalis 6781 Ortho slice#1003	17
3.4 Reduction Issues: An illustrative example of the loss of data resulting from input size reduction for deep-learning. Left: A section of full resolution (2210×1990) Porites Naturalis 6785 annotation, Right: The same annotation after reduction to 224×224	18
3.5 Tiling Approach: An illustrative example of a μ CT slice tiled with window $k \times k$ and step size k	18
3.6 Composite Tile Examples: Example 224×224 tiled image and annotation composites from the Porites Naturalis 6785 dataset.	19
3.7 Volumetric Snippets: An example snippet of dimensions $5 \times 224 \times 224$. Note that the annotation for slice n is in the center of the volume.	19
3.8 Training Augmentations: Some example visualisations of the augmentations applied to the training data. For each image pair the original image is presented left with the resulting transformation presented right. Top Left: Rotation of 180 degrees. Top Right: Rotation of 90 degrees and flip. Bottom Left: Addition of Gaussian noise. Bottom Right: Rotation between $[-20, 20]$ degrees.	20
3.9 Full architecture diagram. (a): A $5 \times 224 \times 224$ input volume is provided to (b): A pre-trained ResNet-50 CNN encoder with 3x skip connections. (c): Feature encodings are blended across the volumetric dimension before begin passed as input to (d): a pre-trained 12 layer ViT with Multi Head Attention and MLP transformer block components (f). (e): Finally, through cascaded upsampling, incorporating the skipped encodings a final segmentation output of size 224×224 is produced (g). Adapted from architecture flow diagram used in [42].	21
3.10 Naive Model Training: Loss and DSC accuracy graphs generated during training of the naive model. Model convergence is reached after ~ 150 epochs with a validation accuracy of 0.59. Note that despite training loss being less than validation, validation accuracy is higher.	22
3.11 Naive Model Inference - Porites Naturalis 6785: Naive model predictions on slices from the same coral colony as the training data. For each pair the left image is the input tile, and the right a composite of the model prediction and ground-truth labels, where red is prediction segmentation and blue is ground-truth. Left Pair: Slice #395: Mean Dice 0.36. Right Pair: Slice #415 (bottom): Mean Dice 0.77.	22
3.12 Naive Model inference - Unknown Colonies: Naive model predictions on slices from different coral colonies of the Porites genus. For each pair the left image is the input tile, and the right a composite of the model prediction and ground-truth labels, where red is prediction segmentation and blue is ground-truth. Left Pair: Porites Astraeiformis: Mean Dice 0.045. Right Pair: Porites Naturalis 6781 Growth slice #1279: Mean Dice 0.1250.	23
3.13 Baseline Training: Loss and Dice-similarity accuracy graphs generated during training of the baseline fine tuned model. Note that convergence occurs very rapidly, achieving a peak validation accuracy of 0.78.	23
3.14 Baseline Inference - Known Colony: Prediction examples from two different test sets of the Naturalis 6781 colony using the baseline fine-tuned model. For each pair the left image is the input tile, and the right a composite of the model prediction and ground-truth labels, where red is prediction segmentation and blue is ground-truth. Left Pair: Growth slice #2499, which is the same projection axis as the training set. Right Pair: Ortho slice #1003, an unseen projection axis of the same colony.	24

3.15 Baseline Inference - Unseen Axis/Colony: Prediction examples from the Ortho and Astraeiformis test sets 3.1.1. For each pair the left image is the input tile, and the right a composite of the model prediction and ground-truth labels, where red is prediction segmentation and blue is ground-truth. Left Pair: Ortho slice #1003, which is the same colony as the training data, but from a different projection axis. Right Pair: Astraeiformis, a completely different species to the training data. Note that the prediction coverage is dramatically improved in comparison to the naive implementation (Fig. 3.12).	24
3.16 Fine-Tuning Training Performance: Composite Loss and DSC accuracy graphs generated during training of fine-tuned models using the Lrg and VLrg training sets respectively. Note that training and validation accuracy's for both fine-tuned models have almost converged. Peak DSC accuracy is noted for both models; Lrg: 0.93, VLrg: 0.97.	25
3.17 Fine-Tuning Inference Comparison - Unseen Axis: Comparative composite ground-truth (blue) overlaid with prediction (red) from the Ortho testing set 3.1.1. From left to right: Baseline, FT-Lrg, FT-VLrg. Note that the VLrg model begins to show better performance at predicting the longer corallite regions which run perpendicular to the projection axis. The black box regions highlight an area in which the corallite region separation improves between models, but which is not necessarily reflected in the DSC similarity scores.	25
3.18 Fine-Tuning Inference Comparison - Unseen Species: Comparative composite ground-truth (blue) overlaid with prediction (red) from the Astraeiformis testing set 3.1.1. From left to right: Baseline, FT-Lrg, FT-VLrg. Note that the larger models are displaying significant issues associated with over-fitting, resulting in many more false-positives where there is background noise in the scan data.	25
3.19 Topological Loss Visualisation: An example of topological error in model predictions, similar to Fig. 2.8, where model segmentations are expressed as connected regions, instead of individual corallites. Topological error map is presented on the left, composite ground-truth (blue) and predictions (red) on the right. The topological loss is designed to proportionally discourage these kinds of prediction errors.	26
3.20 Topological Error: A graphical illustration of the error function $y = \min \left[\left 1 - e^{\frac{A_p - A_l}{A_p}} \right , 1 \right]$ where $x = \frac{A_l}{A_p}$. Note that the rate of change of the error is much more aggressive when $A_p > A_l$, as this is a principle marker of false positive segmentations, which typically manifest as interconnected regions.	27
3.21 Topological Model Training: Composite Loss and DSC Accuracy graphs for three of the topological loss experiments with lambda values of {0.1, 1, 10} respectively. Note the best accuracy/loss values were achieved when the topological lambda factor was 1.	28
3.22 Topological Model Inference - Unseen Axis: Comparative composite ground-truth (blue) and prediction (red) inference outputs from the unseen Ortho testing set 3.1.1. For each pair, the left image is from the FT-VLrg model. The right image is from the same model, with the addition of warmed-up topological loss with $\lambda = 1$. Note the marginal improvement in DSC, but more importantly, the stronger topology in the region shapes, particularly notable around the folded region of the left pair. This will aid reconstruction efforts when generating the 3D model.	28
3.23 Topological Loss + Gaussian Noise Model Training: Composite Loss and DSC Accuracy graphs for three of the Gaussian noise (GN) experiments with topological loss (T) lambda values of {0.1, 1, 10} respectively. Note that the validation accuracy is generally equivalent to the same models without Gaussian noise (3.21).	29
3.24 Gaussian Noise Improvements: Two example composite inference predictions on Porites Astraeiformis data, where blue is ground-truth and red is model prediction. For each pair the left image is predicted by a model trained without Gaussian noise augmentations, and the right with Gaussian noise augmentations. Both models are otherwise identical, having been fine-tuned using the VLrg dataset with 1x topological loss. Note, in particular, the reduction in false positives when Gaussian noise is applied.	29
3.25 Slice Reconstruction Issues: An illustrative example of the appearance of visible seams when reconstructing predictions. Left: A subsection of a prediction reconstruction with tiling step size of 224. Note the red box identifies the regions where visible seams appear, a result of inconsistent model predictions at regions on the edge of the tile. Right: A subsection of the same slice with tiling step size of 112 in which no tile edges are apparent and reconstruction is 'seamless'.	31

3.26	Slice Reconstruction Example: An illustrative example of the use of 224×224 input (yellow) with step size of 112 (red), and the corresponding reconstruction window of 112×112 (blue). This technique is used to generate seamless full-slice segmentation.	31
3.27	Region Axis Properties: A visualisation of the major and minor axis properties inherent in the skimage.measure package. These values can be used with the orientation and center pixel coordinate properties to map ellipses to each of the detected regions at a specific point in space.	32
3.28	Corallite Model Connection: Subset of corallite #27 from the Porites 6781 Growth axis. Left: Instantiation of ellipses in 3D space, with 16 vertices. Right: Resulting corallite structure once regions have been joined.	34
3.29	Corallite Branching: Porites 6781 Growth axis: Corallite #87 (Left) and Corallite #888 (Right) as viewed along the X-axis of the model environment. Note the curvature and branch-like behaviour captured by the implementation.	35
3.30	Colony Modelling: Porites 6781 Growth axis: Partial model of corallites #0 - #20,000 as viewed within model environment	35
4.1	Scan Data Inconsistencies: Left: Composite of prediction (red) and ground-truth (blue) from the Astraeiformis slice. Note the high number of false positive predictions, a result of noise in the background of these slices and insufficient training data for this particular species domain. Right: Porites Naturalis 6781 Growth axis. Note the lack of background noise, which is consistent with the training set, resulting in no false positives in the background region.	38
4.2	Naive Model Inference Limitations: An example set of composite prediction (red) and ground-truth (blue) inference outputs from the naive model on the Naturalis 6785 dataset (3.1.1). Note that although prediction / ground truth overlap is good, and in some cases accurate predictions have been made in lieu of ground-truth labels, a significant number of corallite regions are not captured by this model.	40
4.3	Corallite Cross-Section: An illustration of the apparent corallite geometry (right), depending on the orientation of the corallite growth with respect to the cross-section.	40
4.4	Baseline Inference Examples: An example set of composite prediction (red) and ground-truth (blue) inference outputs from the baseline model. Note that a much broader variety of geometry has been captured by this model, a result of the introduction of the novel annotations. Some exemplary issues with prediction topology from this model have been highlighted by a red box.	41
4.5	Fine Tuning Model Inference Examples:- Astraeiformis: An example set of composite prediction (red) and ground-truth (blue) inference outputs from each of the fine-tuned models on the Astraeiformis slice. Left to Right: Baseline, Fine Tuned Lrg, Fine Tuned VLrg. Note the depreciation in prediction quality is clear, particularly with respect to a significant increase in false positives in the background regions.	41
4.6	Fine Tuning Model Inference Examples:- Naturalis 6781: An example set of composite prediction (red) and ground-truth (blue) inference outputs from each of the fine-tuned models on the Naturalis 6781 Growth slice #2499. Left to Right: Baseline, Fine Tuned Lrg, Fine Tuned VLrg. Note the improvements in DSC metrics, and the topological accuracy illustrated by the regions inside the red boxes.	42
4.7	Topological Loss Inference Examples: An example set of composite prediction (red) and ground-truth (blue) inference output pairs, where for each pair the left prediction is generated by the VLrg model, and the right by the same model with topological loss ($\lambda = 1$). Note the improvement in geometric structure, particularly in the areas highlighted by the red bounding box. Bottom right pair: It is also important to note that topological loss has had little perceivable impact on the false positives generated in the Astraeiformis slice.	43
4.8	Gaussian Inference Examples: An example set of composite prediction (red) and ground-truth (blue) inference output pairs from the Astraeiformis dataset, where for each pair the left prediction is generated by the VLrg model with topological $\lambda = 0.1$, and the right by the same model but with the addition of Gaussian noise augmentations to training data. Note the reduction in false positives around the background regions of the tiles, an issue which had become more exacerbated by previous iterations of the fine-tuning process.	44

4.9	Complete Slice Segmentation's: A randomly selected set of segmentation slices from the Naturalis 6781 colony scans, generated by the optimal model identified by this thesis, outlined in section 4.2.1. (1): Growth slice #1919, (2): Ortho slice #2368, (3): Growth slice #1194, (4): Ortho slice #2078. Note that, in general, the model achieves excellent separation between corallite regions, and is able to describe a range of region geometry representative of the omni-directional corallite growth patterns. This is crucial for the success of the corallite reconstruction phase of this project.	47
4.10	Model Connectivity Behaviours: Illustrative examples of adversity in corallite connectivity behaviours in relation to thresholds of IoU and L2 norm when tracing corallite paths. Left: Threshold values are too high, resulting in separate corallite structures, where one should be formed as two corallites merge. Right: Threshold values are too low, resulting in one corallite structure, where there should be several.	50
4.11	Missed Regions: An illustrative example of regions missed by the segmentation model resulting in separate corallite structures. When a region at height n is unable to find a region at height $n - 1$, which meets the parameters of the search described in section 3.3.2, a new corallite structure is formed. This example shows five 'unique' corallites, where there should be one. Extending the depth of the search for region-corallite mappings would help to further mitigate these issues.	51

List of Tables

3.1	Raw Data Summary: Summary table of the raw data used for this project provided courtesy of the 4D-Reef Project, The Natural History Museum and the University of Bristol School of Earth Sciences. The strength of the annotation quality refers to completeness of labelling, as discussed in section 3.1.2. Dataset colour coding for ease of reference to the annotation details set out in Table 3.2.	14
3.2	Annotation Details: Detailed breakdown of the properties of each of the novel annotated slices provided by this project. The circular and elliptical regions are approximate counts. A region is considered circular if the absolute difference between the lengths of the major and minor axes is less than 10 pixels. An illustration of these axes are presented in Fig 3.27. Colour coding indicates the colony dataset from which each slice was drawn, according to descriptions provided in Table 3.1.	15
3.3	Data Separation: Separation of datasets for testing and training for both the naive model, and for later, fine-tuned models. The naive model is trained and tested on the same dataset, the weakly annotated Naturalis (Nat.) 6785, separated by an 80/20 split with no overlap. The fine-tuned models are pre-trained on 100% of the Naturalis 6785 data, and then fine-tuned on one of the complete Naturalis 6781 slices: #1279 from the Growth axis projection. Testing sets for the fine tuned models include: (1): Naturalis slice #2499, also from the Growth projection axis, but well separated by distance within the scan volume. (2): Slice #1003 from the Ortho projection axis, unknown to the trained models. (3): A slice from the Astraeiformis species of the Porites genus, unknown to the trained models, with completely unknown scan parameters, position in the colony, and axis of projection.	15
4.1	Naive Model: Performance of the Naive model on the Naturalis 6785 testing set. The Naturalis 6785 data is unsuitable for performance analysis of fine-tuned models due to the sparsity of annotations, however the results are indicative of performance when applied to the same data domain as the training set. DSC: Mean Dice similarity, Topo: Topological accuracy as defined in 2.2.4.	48
4.2	Inference Results: Full testing performance for all model iterations. The model in bold is selected as it offers the best, or joint best, performance across all testing domains, with the exception of the topological accuracy on the Astraeiformis data. Naturalis Growth #2499 is an unseen slice from the same colony and same projection axis as the training data. Naturalis Ortho #1003 is an unseen slice from the same colony but different projection axis as the training data. Astraeiformis is an unseen slice from a different colony, and species, of the Porites genus. α Topo: The lambda factor applied to the topological loss function. (P): Phase-in of topological loss over 100 epochs. + GN: Addition of Gaussian noise to training data. DSC: Mean Dice similarity, Topo: Topological accuracy as defined in 2.2.4.	48
4.3	Alternative Topological Accuracy Metrics: Alternative metrics used to support the hypothesis that topological loss improves topological accuracy. Testing was conducted on the Naturalis 6781 Ortho dataset 3.1. Compactness: Measures how ‘clean’ the geometric structure of each region is. A lower perimeter-to-area ratio suggests tighter geometry, which is preferable for the region-fitting tasks when generating the three-dimensional reconstruction. Corallite Count Error: The error in total number of corallites predicted in a complete slice. Unique Corallite Count: The total number of unique corallites counted in the final full-colony model.	49

Ethics Statement

This project did not require ethical review, as determined by my supervisor, Dr. Tilo Burghardt. Ethics in relation to coral data acquisition are covered by the 4D-Reef project¹, who provided the data for this thesis.

¹<https://www.4d-reef.eu/>

Notation and Acronyms

The following list provides a quick-reference resource for acronyms and notations used throughout this thesis.

(μ)CT	:	(Micro) Computed Tomography
DL	:	Deep Learning
RGB	:	Red, Green, Blue
ANN	:	Artificial Neural Network
CNN	:	Convolution Neural Network
ViT	:	Vision Transformer
TCM	:	Temporal Context Module
BCE	:	Binary Cross Entropy
(R)MSE	:	(Root) Mean Squared Error
IoU	:	Intersection over Union
SDPA	:	Scaled Dot-Product Attention
MHA	:	Multi-Head Attention
SotA	:	State of the Art

Chapter 1

Introduction

This thesis introduces a data-driven approach to 3D individual corallite reconstruction, using volumetric vision transformers and topological loss. Building upon leading work in the medical imaging domain, and supported by ablation study, a state-of-the-art segmentation solution for individual corallite detection is developed to facilitate the automated generation of a fully interactive 3D representation of each unique corallite of a given colony. To the knowledge of the author and his supervisor, this is the first example of such a system and it is developed in the hope that it can be used to further understanding of coral polyp development, and the influences of the marine environment on one of our worlds most valuable ecosystems.

Context: The order classification of Scleractinia, or stony corals, comprise colonies of small organisms called polyps. These polyps secrete Calcium Carbonate ($CaCO_3$) in the form of a protective cup, known as a corallite, creating a lasting skeleton of the structure of the colony [34]. The Porites genus of Scleractinia are found in oceans and seas throughout the world, with living colonies of over 400 years old [36], offering a representative general domain in which to explore this thesis. As the colony develops, polyps are formed and reabsorbed in accordance to access to resources in localised regions [1]. Much like the rings of a tree, the influences of the environmental past can therefore be understood through the patterns of corallite growth within the colony skeleton. Understanding these changes offers insights into the impacts of environmental factors on the overall growth and success of colonies, and thus how we might better protect these essential coastal organisms [25].

Image-based deep learning techniques present an opportunity to better understand coral life cycles by expediting the process of building a spatial-temporal model of corallite growth. Presently, to track the development of a given corallite and by proxy, the polyp which formed it, painstaking manual analysis of the colony skeleton and μ CT scan data is required. Further, expertise is necessary to translate this analysis into a useful three-dimensional visual representation. This project sets out to show that a deep-learning model can predict, with reasonable accuracy, corallite growth across the entire colony and provide suitable data to generate a model for three-dimensional visual analysis and better spatial-temporal understanding of corallite growth.

1.1 Goals and Objectives

One of the fundamental goals for this project is to produce a model of the trajectory and shape of each unique corallite through the extent of the coral colony. In order for the pipeline to successfully build a 3D representation of each of corallite, it is necessary that the model has reliable geometric structure and proper separation between unique corallites. This is a non-trivial task as colonies typically consist of many thousands of corallites, corallite growth is not linear, and a corallite may split or join at arbitrary points throughout the lifespan of the colony [7],[6].

This project can be viewed in three key phases, describing a viable system for volumetric prediction and reconstruction. Each phase forms the basis of organisation for the discussion of this thesis, and is outlined in brief below, noting important research questions or technical requirements which this project will aim to address.

Novel Volumetric Data Generation: The first objective of this thesis is to explore how μ CT data of coral colony skeletons can be leveraged to develop a automated corallite reconstruction pipeline. The raw μ CT data requires annotation and reshaping, for training a deep-learning segmentation model to predict corallite regions. Corallite growth is omni-directional and, if this task is to be successful, there are important relationships to capture in all three dimensions. This section of the thesis will focus on some of the opportunities and limitations of the μ CT data by considering the quality of raw data and any associated labelling, particularly with respect to its coverage of the domain and how effective it is in supporting a general corallite detection solution. Further, the preparation and implementation of volumetric-based data structures will be explored for use in a state-of-the-art deep learning solution.

Deep-Learning Pipeline: The development of state-of-the-art volumetric segmentation models will be explored in this section, expanding on work done in [5] and [42]. Specifically, the pipeline will encode the relationships of the two-dimensional slice data across a third dimension, representing the projection axis of the scan. The deep learning (DL) model pipeline leverages the benefits of both UNet and Vision Transformer architectures. Through its development, this section of thesis sets out to understand the following questions:

- **DL 1:** To what extent does a volumetric segmentation model, designed for μ CT data in the medical imaging domain (eg. VT-UNet [42]), translate to the domain of μ CT corallite segmentation?
- **DL 2.1:** To what extent does the introduction of a custom loss function improve on a baseline state-of-the-art segmentation solution?
- **DL 2.2:** To what extent does the introduction of a custom loss function support the generation of a 3D representation of unique corallites?

Experiments will support the identification of the most suitable model parameters, and detailed quantitative and qualitative evaluations of the models strengths and weaknesses will be discussed.

Corallite Reconstruction: The final phase of this project explores prototyping the automated generation of a useful three-dimensional reconstruction of all the corallite regions within a colony skeleton. This section of the thesis will explore the use of open-source modelling software Blender¹ to translate the model outputs into geometric object data which can be read by all contemporary modelling systems. One of the key opportunities such a system presents is the ability to isolate a model of an individual corallite, with viewing capabilities in all three degrees of freedom. This will facilitate detailed analysis of corallite formation throughout the colony, without the need to manually trace development through the μ CT scan data. Given the time constraints of the project, the opportunities and limitations of one possible approach to 3D corallite reconstruction will be considered, with particular importance placed on capturing the non-linear growth and the branch-like behaviour of corallite development.

1.2 Contributions

The work done in this thesis can be viewed as a significant step towards fully-interactive spatial-temporal reconstruction of individual corallites from colonies of the *Porites* genus. Throughout it's exploration, the following summary contributions are made:

- The creation of a novel set of corallite region annotations, consisting over 8,400 labels, from micro-computed tomography (μ CT) data of multiple colonies of the *Porites* genus. The raw μ CT was provided courtesy of the University of Bristol Earth Sciences department, as part of the 4D-Reef project².
- The implementation of a state-of-the-art volumetric-based binary segmentation pipeline, adapted from the ViT-UNet architecture proposed in [42].
- The development and evaluation of a new, novel, custom topological loss function to support domain specific corallite region mapping with reliable geometric structure.
- The implementation of a novel prototype approach to automating 3D individual corallite reconstruction using volumetric segmentation data.
- Qualitative and quantitative evaluation of all key stages of the proposed corallite reconstruction pipeline, on two complete coral colony scans, plus additional supplementary μ CT from two other separate colonies.
- Proposals for a broad range of future research opportunities which may serve to further enhance the work in this thesis, and the study of polyp development as a whole.

¹Blender: <https://www.blender.org/>

²4D-Reef: <https://www.4d-reef.eu/>

Chapter 2

Background

This chapter provides a contextual background to the domain of coral study, and pre-requisite technical knowledge for the implementation used by this thesis. A brief application background will summarise the context from an environmental and geological sciences perspective. The technical background will follow, in which an initial overview of the image data acquisition and representation precedes a thorough coverage of deep learning models, with particular attention given to architectures suitable to the vision-based tasks relevant to this problem. Finally, an overview of related previous works will be provided.

2.1 Coral Ecology

This section will provide a brief overview of the necessary background from an coral ecology perspective. It will focus on the importance of coral reefs within our marine ecosystems, the order of Scleractinia, and finally the importance of understanding corallite growth within both of these contexts.

Coral Reefs: Coral reefs are among the most densely populated areas of biodiversity in our oceans. They serve as a critical component of the oceanic ecosystem by providing nutrition and a safe haven for a wide range of species [38]. They protect coastlines from storms and erosion, are a source of food and medicines [26], and generate an estimated \$30bn USD for the global economy [4]. However, climate change poses a significant threat to their survival. Warming and acidification of seas and oceans have increased the risk of coral extinction within this century [16]. It is necessary to therefore build a detailed understanding of the factors affecting the success of coral colonies, in order to better protect this crucial environment.

Scleractinia and Porites: The order of Scleractinia, or stony corals, comprises over 1400 different species of coral, many of which are reef-building, shallow water corals, such as the Porites genus, and are considered to be the 'architects' of our oceans coral reefs [20]. Fossils of Scleractinia date back to the middle Triassic period, and have therefore been a feature of our oceanic ecosystems for more than 230 million years. The Porites genus of Scleractinia are found throughout the world and form living colonies of over 400 years old [36]. They are considered to be one of the key reef-framework builders [28], and therefore serve as an excellent case-study for investigation.

Importance of Corallites: Stony corals have been shown to form by division through asexual reproduction, known as budding, from a founding polyp [6]. Polyps produce a protective cup-like structure of calcium carbonate ($CaCO_3$) known as a corallite, mm in thickness, which thus form the basis of a colony skeleton [7]. The number of active polyps within a Porites colony is dependent on its size, however with corallite diameters of up to only 2mm, these numbers can exceed tens, to hundreds of thousands [28]. Some example imagery of corallite details are presented in Fig. 2.1.

Manual analysis and 3D visualisations of coral colonies have suggested evidence of dynamic behaviours in polyp budding [23], resulting in complex branching and merging within the corallite structures. These findings led to work which considered the demographic dynamics of polyps [2], showing that polyps in central regions of the colony tend to live longer than those in outer and lateral regions. This work further suggests that colonies undergo complete turnovers of polyp populations throughout its lifespan, at an estimated rate of eight times per hundred years.

Inspired by these findings, further research has shown that colony growth is a direct result of the extension and division of polyps, and that the budding of new polyps is dependent on location and access to resources, rather than age [25].

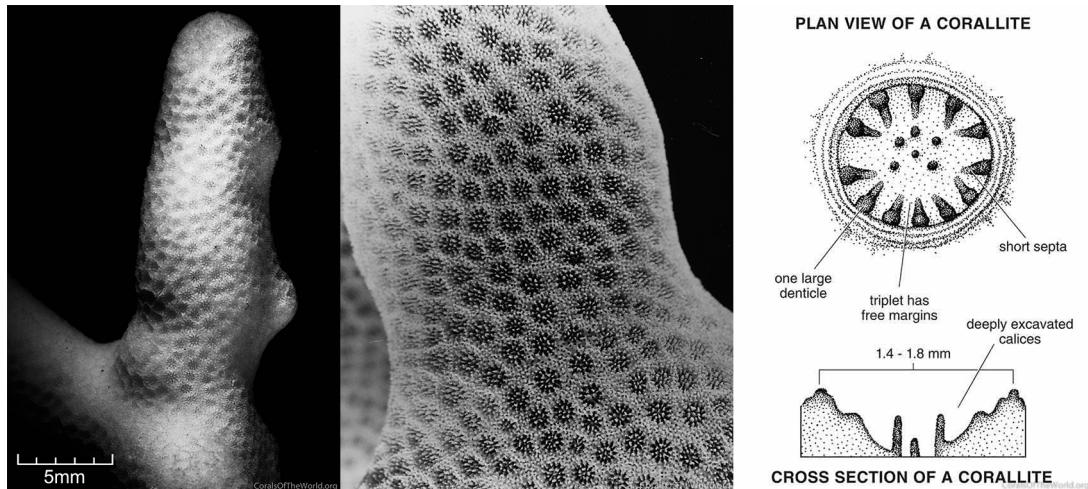


Figure 2.1: **Corallites:** As selection of images presenting corallites from a variety of perspectives. **Left:** A section of a *Porites Compressa* colony. **Center:** Close-up of the same section illustrating the distribution and density of the corallites. **Right:** Diagram of the structure of a single corallite. Note the size of the corallite region is 1.4 - 1.8mm only. Image credits: [Corals of the World](#).

Summary: In summary, much research has shown that polyp development, and the resulting colony skeleton, is not only highly complex, but encodes important information about the condition of the colony, its environment, and its overall health at the time the calcium carbonate formation occurred. In combination with historic data of the oceanic environment, such as temperature and pH, a detailed understanding can be formed of how a coral responds to environmental factors. This motivates the importance of accurately modelling the development of corallites, and advances in μ CT scanning technology have provided the data which is detailed enough to facilitate this.

2.2 Technical Background

This project will consider the task of corallite predictions as a volumetric binary segmentation problem. The key goal is therefore to classify each voxel (a 3D pixel) of the input data as to whether or not it belongs to a corallite region. The project will take a deep-learning approach using sequential μ CT slice data to form the volumetric data. This section defines the format of the underlying data and associated labels, before covering the fundamental deep-learning components leveraged by the solution.

2.2.1 Image Representation

The image data used throughout this project is provided courtesy of the 4D-Reef Project¹, The Natural History Museum² and the University of Bristol School of Earth Sciences. The data consists of cross section ‘slices’ from a variety of coral colony skeletons of the *Porites* genus. Each slice has been constructed from micro-computed tomography (μ CT) data of the skeleton and when stacked along the projection axis, they form a three-dimensional representation of the coral. Figure 2.2 illustrates the cross-sectional intersection of these slices.

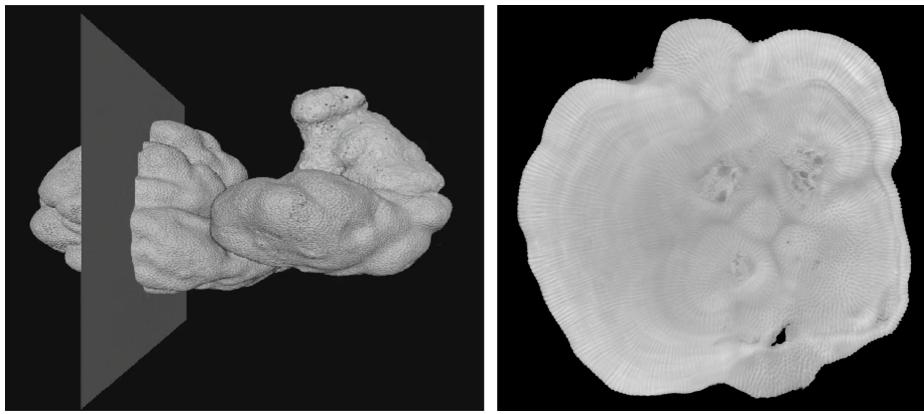


Figure 2.2: μ CT Slices: An illustrative representation of how a cross-sectional slice is taken from μ CT data of the full colony skeleton. The projection axis, throughout the context of the discussion in this thesis, is perpendicular to the plane formed by the slice. Image credit [32].

The properties of the slice data varies from colony to colony as the μ CT equipment and capture techniques differ. As a result of the second-hand nature of the data, mixed information is available about the capture parameters but invariably the image resolution is a minimum of 1990×2200 pixels, in single channel gray-scale colour representation. Formally, for each image I :

$$\forall x, y \in I : 0 \leq I_{x,y} \leq 255$$

Annotation Data: In addition to the raw image data, a subset of sequential corallite region annotations were provided by the University of Bristol Earth Sciences department, created for use on similar deep learning tasks. These annotations were generated using the GNU Image Manipulation Program³ and are provided as three channel (RGB) image data, with pixels corresponding to a corallite region labelled [24, 28, 37] respectively. The colour choice is arbitrary, and given the segmentation task is binary, these annotations are transformed to single channel gray-scale representation I' using the Python OpenCV library⁴ with the following conditions, where c is the RGB channel with $|c| = 3$:

$$\forall x, y \in I : I'_{x,y} = \begin{cases} 255, & \sum_i I_{x,y,c_i} > 0 \\ 0, & \text{otherwise} \end{cases}$$

Finally, four additional slices have been annotated specifically for this project using Adobe Photoshop⁵. The annotation data is provided in single-channel gray-scale format according to the aforementioned conditions and are presented alongside the respective raw inputs in Fig. 3.3.

¹4D-Reef Project: <https://www.4d-reef.eu/>

²NHM: <https://www.nhm.ac.uk/>

³GIMP: <https://www.gimp.org>

⁴OpenCV: <https://opencv.org>

⁵Adobe Photoshop: <https://www.adobe.com/uk/products/photoshop.html>

2.2.2 Deep Learning Architectures

Machine Learning is concerned with the study of applying statistical mathematics to build predictive models which improve in accuracy over time, and as more data is provided [29]. Machine learning solutions can be broadly categorised into one of three groups, related to the learning approach taken: supervised learning, unsupervised learning, and re-enforcement learning [3]. This project will take a supervised learning approach. In the supervised learning paradigm, models are *trained* according to some pre-defined error metric which measures similarity between the model predictions and manually generated target outputs [3].

Deep learning specifically is a subset of machine learning concerned with Artificial Neural Networks (ANNs) of depth > 3 , defined in [9] as:

"... a class of machine learning algorithms that: (1) use a cascade of multiple layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input, (2) learn multiple levels of representations that correspond to different levels of abstraction; the levels form a hierarchy of concepts."

This section will provide a brief overview of ANNs in general, before covering in more detail the fundamental building blocks used in vision-based deep learning, specifically those employed in the architecture used for this project.

Artificial Neural Networks (ANNs): The McCulloch-Pitts neuron [24] was the first computational model of a neuronal activity within a network. Its key observation is the propagation of a signal through the network based on whether a weighted sum of input signals meets some threshold. Formally, for each node i , the output y_i is the result of a threshold function Θ , where x_{ij} is input from node j connected to i , and w_{ij} its corresponding weight.

$$y_i = \Theta\left(\sum_j x_{ij} w_{ij}\right)$$

Using this simple proposition, complicated multi-layered *feed-forward* networks, also known as Multi-Layered Perceptrons (MLPs), can be constructed (Fig 2.3). Due to the typically dense connectivity of these networks, high dimensional information can be learned by encoding the weights associated with each node connection [3]. Central to this learning is the concept of back-propagation, first introduced in the context of ANNs by Werbos in the 1980s [39]. In simple terms, once an input has passed through the network, an error signal is generated in the form of some loss function between the output of the model and a target. Back-propagation is the process of passing this error back through the layers of the network by differentiable adjustment of the weights. The input signal is once more submitted to the network, generating a new error. This cycle repeats in what is commonly known as *epochs* until some point of convergence is reached, ie. a minima of the loss function.

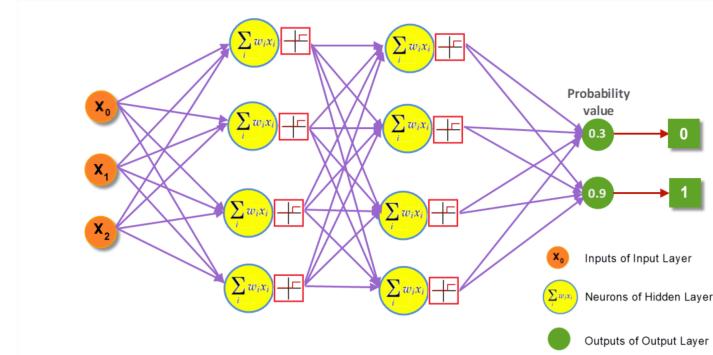


Figure 2.3: **Artificial Neural Network:** A diagram of a fully connected artificial neural network. Inputs $x_i \in \mathbf{X}$ are fed through two hidden layers consisting activation functions based on a weighted sum of the inputs. The activation is triggered if and only if that weighted sum passes some threshold. The output layers represent probabilities that the input belong to each of the N classes, in this case $N = 2$. Image credit: [DevSkrol](#).

Convolutional Neural Networks (CNNs): When it comes to image-based deep learning tasks, one of the challenges presented by a traditional feed-forward network is each pixel is treated as a singular unique input value. Even for very low resolution inputs, this property demands large, complex networks. Further, the image is necessarily flattened prior to input, resulting in the loss of pixel-to-pixel contextual information. Inspired by models of the visual cortex, in which neurons respond to a small surrounding stimulus region known as the receptive field [13], early CNNs were developed to both encode contextual information and reduce dimensionality through the use of convolution kernels [14], [21].

Convolution Kernels: A convolution is an arithmetic matrix operation performed in many classical image processing techniques. It involves the use of a small $n \times n$ kernel, which is stepped across the image data in what is commonly referred to as a *sliding window* sequence. At each step an element-wise operation is performed between the values within the window frame and the kernel itself. A simple example is presented in Fig. 2.4 in which the output C is the sum of the element-wise multiplication of a 3×3 kernel K sliding across the input image I , formally:

$$C_{x,y} = \sum_u \sum_v K_{u,v} I_{x+u, y+v}$$

Note that the output encodes a relationship between each of the elements in the window at a given step. It is also important to note that convolution kernels can be used to reduce dimensionality, as in the example presented in Fig. 2.4, but padding techniques can be employed if it is desirable to maintain dimensionality. In traditional image processing techniques the values of the kernel would be designed 'by hand' to extract the desired features, for example, using the Sobel operator to detect edges within an image [17]. In the domain of CNN's the kernel weights are one of the key parameters to be learned by the model as part of the loss-minimisation process.

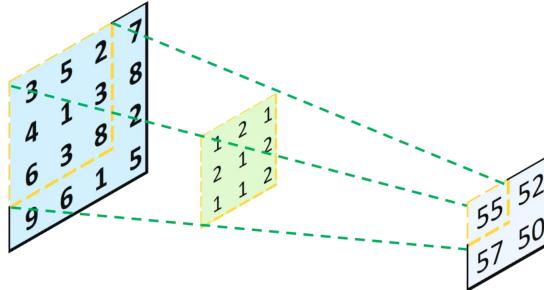


Figure 2.4: **Convolution Kernels:** A toy convolution kernel operation in which the output is the sum of the element-wise multiplication of the kernel and its window. Image credit: [Towards Data Science](#).

Pooling: Pooling is an important step in convolution-based deep-learning architectures as it is frequently used as the key operation in which dimensionality reduction occurs. There are a variety of different pooling techniques, but work in [33] has shown max pooling to be most effective, at least for object recognition tasks. The max pooling operation is a process of sliding a fixed sized $m \times m$ patch over the input, with step size m , and taking the maximum value of the input space currently observed within the patch, as described in Fig. 2.5.

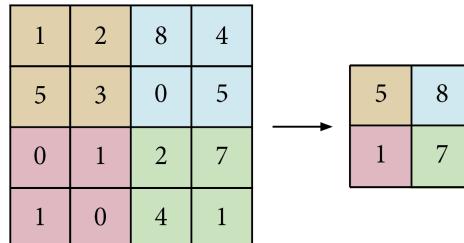


Figure 2.5: **Max Pooling:** An toy example of a max pooling operation with patch size 2×2 . Note the reduction of each patch is simply the maximum value within that section.

Up-sampling: Following a series of convolution and pooling operations, the input data has been encoded to a lower dimension. In many cases, it is desirable for the final output of the model to have the same dimensional properties as the original input, for example, in the form of a pixel-wise image segmentation map. Therefore it is necessary to 'decode' the data via a process of up-sampling. This project will use two-dimensional bi-linear up-sampling, a frequently used technique which inserts interpolated data points between existing ones, as shown in Fig. 2.6.

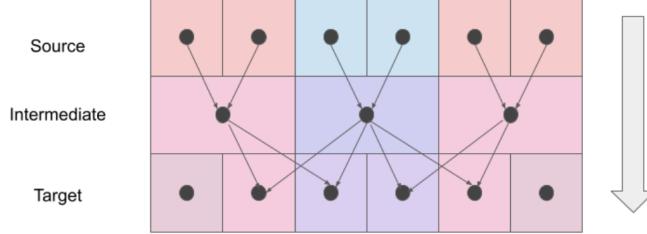


Figure 2.6: **Bi-Linear Sampling:** This image depicts bi-linear down-sampling from Source to Intermediate, followed by bi-linear up-sampling from Intermediate to Target. The lines represent contributions to the given interpolated element in the direction of the down-facing arrow. Image credit: [Bart Wronski](#).

UNet Architecture: The UNet architecture was first introduced in 2015 for biomedical image segmentation [31], and has had significant impact on contemporary segmentation efforts since. The etymology stems from the shape of the architecture, observed in Fig. 2.7, which can be broadly separated in three main components; encoder (left), bottleneck (center), and decoder (right). The UNet architecture is a general design framework and the properties of each of these three components differ between implementations. The model is typically trained at all three stages to output a pixel-wise segmentation map in which each pixel is classified as belonging to 1 of N classes. It is common, as is the case in this project, to use pre-trained components for one or more stages within a UNet architecture. Typically, these will have been trained on vast amounts of data and have been shown to generalise well when fine-tuned for the task at hand [15].

The encoding section down-samples the input through a combination of convolution layers and pooling. The bottle-neck is where most architectural variation exists between UNet-style models but in its simplest form, might be an MLP which takes n encoded inputs and produces m outputs. Finally, the decoder section takes the outputs of the bottleneck and generates a pixel-wise segmentation map through a process of up-sampling.

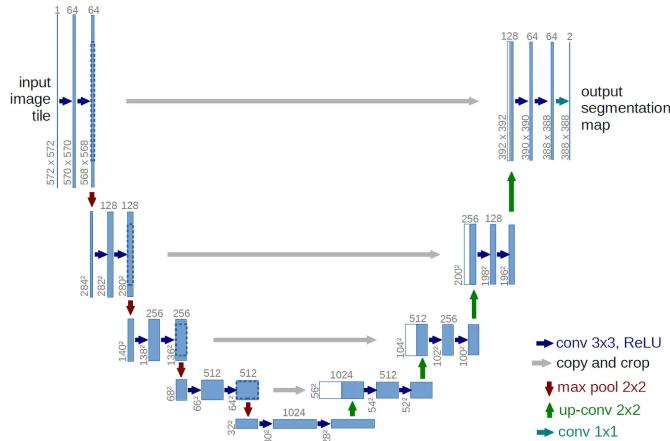


Figure 2.7: **UNet Architecture:** An example UNet architecture. Encoding occurs on the left side of the architecture, where the input is down-sampled. The bottle-neck is the central section. Decoding occurs on the right side when the outputs of the bottle-neck are up-sampled to the original input size. The grey horizontal arrows represent *skip connections*, discussed in section 2.2.6. Image credit [Towards Data Science](#).

2.2.3 Loss Functions

Loss functions are a central component to a deep learning model as they define the landscape to which the model will be trained to find an optimal, preferably global, minimum. In order to find these minima a variety of algorithms can be used, and many deep-learning optimisation techniques target making this landscape easier to navigate [22]. In a supervised learning context, the loss function typically defines some measurement of success between the target and the model predictions [3]. This project will use a combination of Binary Cross Entropy (BCE) and the Dice Similarity Coefficient (DSC) as a baseline loss function, followed by the introduction of a custom topological-based loss to investigate how topology might be used to better define the structure of the segmentation output.

Binary Cross Entropy Loss (BCE): The term *entropy* has roots in information theory and is related to the number of bits required to transmit a random event from a probability distribution [35]. In simple terms, it describes a measure of uncertainty given a proposed distribution, thus *cross-entropy* describes a measure of uncertainty between two-distributions. In the supervised learning context, these two distributions are the target and the predictions. Formally BCE is described by the following equation:

$$\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^N y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i))$$

Sørensen–Dice Coefficient: The Sørensen–Dice coefficient, or Dice Similarity Coefficient (DSC) is a measure of similarity between two sets of data points by the following equation:

$$DSC = \frac{2|A \cap B|}{|A||B|}$$

This computes the ratio between the intersection of the two sets and the product of their respective cardinalities. The Dice similarity is commonly used in vision-related deep learning models as it is quick and easy to compute, is able to compare sets of different sizes, and gives normalised outputs in the range of $\{0, 1\}$. Because we are interested in minimising the loss with respect to the DSC we define the Dice Loss as:

$$\mathcal{L}_{DSC} = 1 - DSC$$

Topological Loss: This project introduces a custom topological loss function, designed to pay attention to regions of the predictions which show significant topological error. In the context of corallites, this typically manifests as incorrectly connected region predictions, identifiable in the images presented in Fig. 2.8, where blue is ground-truth label and red is prediction. Note the red segmentation forming connections between the correctly predicted (purple = blue + red) corallite regions. Implementation details are discussed in section 3.2.3.

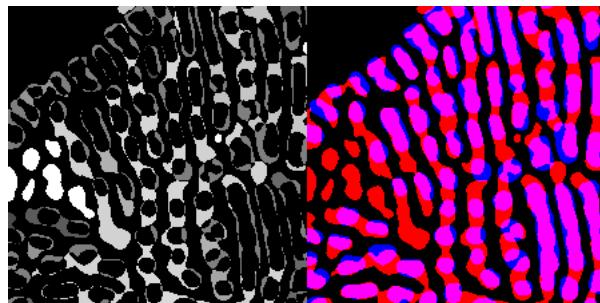


Figure 2.8: **Topological Loss Visualisation:** An example topological error map (left) where brightness indicates larger error. Corresponding prediction/ground-truth overlay (right) where predictions are red, ground truth blue and therefore, correct pixel-wise predictions present as purple. Note that the errors are of a topological nature, a result of false positive predictions forming connections between corallite regions.

2.2.4 Performance Metrics

Closely aligned with the loss function, is the importance of having suitable metrics to measure the performance of the model in the context of the task. In many cases this might simply be directly related to the loss function, ie. $1 - \mathcal{L}$.

This project predominantly uses the Dice Similarity Coefficient described in Section 2.2.3 to measure image accuracy during training and validation phases, and later introduces a topological accuracy measure. During testing, RMSE will be used to measure the accuracy of the corallite region center points, as an additional means to understand the broader performance of the models.

Root Mean Square Error (RMSE): RMSE measures the distance between a vector of predictions and a vector of targets. In terms of corallite center accuracy analysis, each prediction center p in \mathcal{R}^2 is compared with its nearest neighbour ground-truth center l_{NN} . Formally:

$$RMSE = \sqrt{\sum_i^N \frac{(pc_i - lc_{NN})^2}{N}}$$

Topological Accuracy: The topological accuracy metric is used as a measure to understand the topological loss using the pixel-wise error map \mathcal{E} , detailed in Section 3.2.3. Formally:

$$\mathcal{A}_{TOPO} = 1 - \frac{1}{N} \sum_x \sum_y \mathcal{E}_{x,y}$$

2.2.5 Transformers

The seminal paper Attention Is All You Need [37] introduced the Transformer model to the field of deep learning through the context of machine language translation. Transformers have since been widely adopted and have shown to have a profound impact across a broad range of machine learning tasks, including machine vision [18]. The encoder and decoder components of a Transformer model are a construction of N identical layers (Fig. 2.9) and by use of the multi-head attention mechanisms within these components, global and sequential relationships can be built into the model.

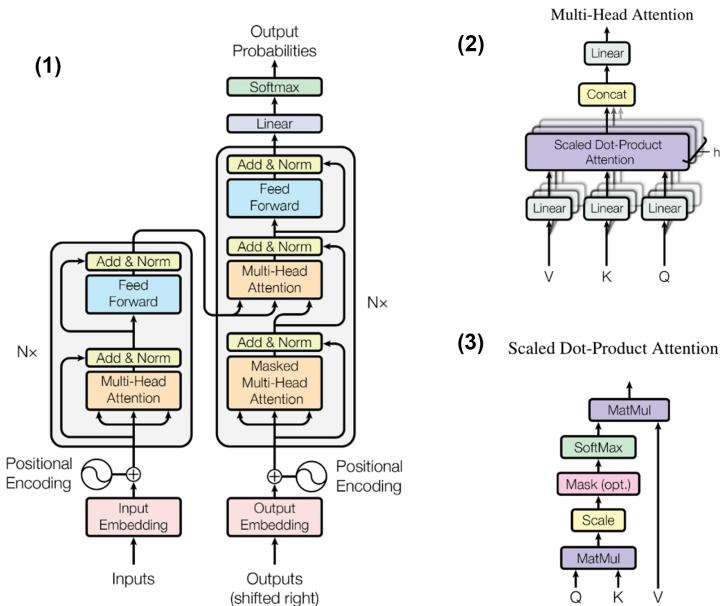


Figure 2.9: **Transformer Component Architecture:** (1) Transformer encoder/decoder components. (2) Multi-Head Attention (MHA) component. (3) Scaled Dot-Product Attention (SDPA) component, as described in [37]. The MHA provides contextual encoding through concatenation and linear projection of a number of sequential SDPA calculations, where Q, K and V represent Query, Key and Value respectively.

Vision Transformers (ViTs): One of the limitations of CNNs is that by nature of the size of the convolution operations, the contextual information in the resulting encoding is limited in range [18]. The Vision Transformer (ViT) was developed to alleviate this issue by adapting the traditional Transformer model to use a linear projection of flattened input patches of the image as a token sequence (Fig. 2.10). This approach leverages the Transformers attention mechanisms to encode relationships and dependencies from the entire input sequence. Vision Transformers have been shown to exceed performance of state of the art UNet CNN approaches for image classification tasks [10], but require pre-training on vast amounts of data in order to compensate for the inductive biases usually formed by CNNs.

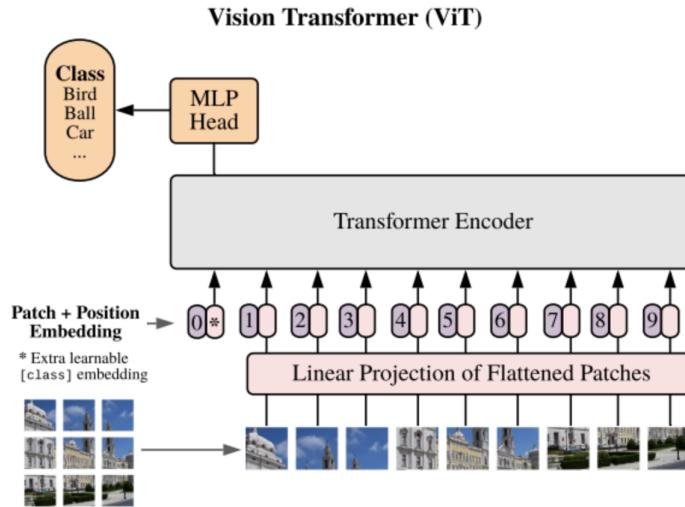


Figure 2.10: **Vision Transformer Architecture:** An illustrative example of the ViT architecture used in [10]. Note the input is separated into a sequence of patches, and associated positional embeddings. These patches are flattened by linear projection before being fed into the traditional transformer network architecture.

2.2.6 Deep Learning Optimisation Techniques

Effectively training a machine learning model can be extremely costly in both time and money. There are a number of properties within deep-learning architectures, often referred to as *hyper-parameters*, which are a common focus for machine-learning optimisations. This sub-section will provide an overview of some of the key techniques explored in this thesis as part of developing a state-of-the-art deep learning pipeline for corallite segmentation.

Learning Rate: Learning rate is a crucial hyper-parameter used in deep learning to control the extent to which a model adjusts its weights, given the gradients calculated during back-propagation [41]. Typically, the learning rate is kept low to avoid missing the minima when performing the gradient descent process, however this can slow down the process of convergence as the step-size is essentially reduced [40]. Modern models employ the use of specific optimisation algorithms to give dynamics to the learning rate throughout the training process, such that it is generally lowered as the model nears convergence.

Optimisers: A variety of optimisers have been developed to improve deep-learning model training. This project will use the Adaptive Moment Estimation (Adam) optimiser. One of the key benefits of the Adam optimiser is that it applies an adaptive learning rate to *each* parameter based on the first and second moments of the respective gradients [19]. This greatly improves efficiency and removes / reduces the need to employ manual learning rate tuning procedures.

Transfer-Learning, Fine-Tuning & Layer Freezing: There exists a wealth of publicly available models, pre-trained on huge sets of data, which have been shown to have excellent general performance within their domain [11]. Transfer-learning is the process of adapting these well-trained 'general' models to a specific task, through processes such as fine-tuning and layer freezing. Transfer learning has been shown to be extremely effective across a broad range of tasks, and has become an important part of

contemporary deep-learning practise [12].

Layer freezing is the process of declaring that the gradients of particular components of a deep-learning network should not be updated during the back-propagation optimisation process. In the context of fine-tuning, the weights of the pre-trained model are retained while the gradients deeper into the pipeline and more relevant to the target application are updated to support task-specific predictions. This project will leverage transfer-learning techniques to build a baseline model for predicting corallite regions volumetrically using a ResNet-50 backbone and the weakly annotated Naturalis 6785 dataset, before fine-tuning on the novel, strongly annotated Porites 6781 dataset.

Skip-Connections: A side effect of the dimensionality reduction within CNN architectures is a loss of some of the finer features within the input data. Skip-connections have become an important part of contemporary CNN design, commonly used in architectures such as Residual Net (ResNet) [15] as they support the preservation of information at each layer of the convolution process by passing it directly to the decoder section of the architecture. An additional benefit of skip-connections is that the back-propagation process is greatly simplified at these steps, improving computational efficiency, and therefore training time. [22] illustrates the impacts of skip connections on the loss-landscape of ResNet56, as depicted in Fig. 2.11.

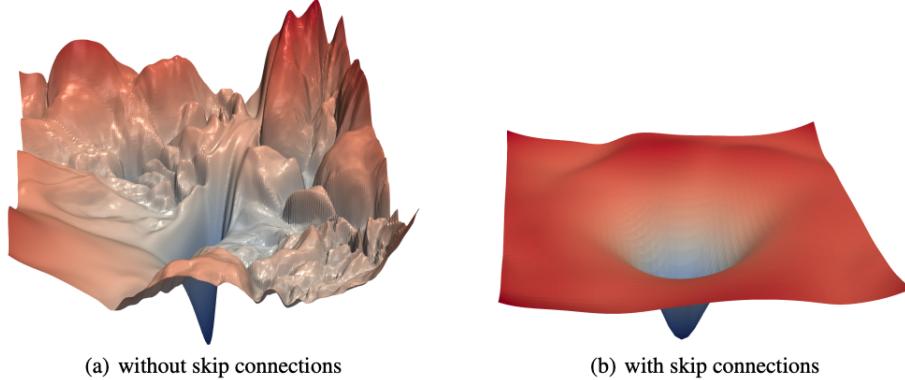


Figure 2.11: **Skip Connections and Loss-landscape Visualisations:** The impacts of skip-connections on the loss landscapes of ResNet56, visualised by [22]. Note the how the resulting landscape is much smoother, simplifying the process of finding a global minimum.

Warm-Up & Phase-In: The warm-up technique is commonly employed for dynamically adjusting model parameters. For instance, in [15], warm-up is used to refer to the initial phase of training during which the model uses a reduced learning rate until the loss falls below a certain threshold. By doing so, the model is gradually brought closer to convergence before the learning rate is increased.

An extension of warm-up is the phase-in technique, which involves gradually adjusting a parameter between its initial and target values after the warm-up period, rather than changing it abruptly. In this project, the use of warm-up and phase-in will be considered when introducing a custom loss function as an additional measure after an initial period of training.

2.3 Previous Work

This section will introduce three previous works relevant to the architecture used for this project. The first, an unpublished work, specifically in the domain of corallite prediction, uses object detection techniques to find corallite region center points, using approximate bounding boxes to define the regions. By considering this a segmentation problem, this project takes an alternate approach, as it views the region topology of equal importance to the centers when segmenting for three-dimensional model generation. More closely aligned with the approach this project takes, Trans-UNet [5] considers the application of a UNet-Transformer-based architecture for image segmentation tasks, and Video Trans-UNet [42] expands on this model to incorporate a Temporary Context Module (TCM), encoding sequential video frames into the network.

Corallite Center Detection: Work done in a previous masters thesis sought to address the problem of corallite region prediction by leveraging the then state-of-the-art YOLO v3 [30] object detection framework. The project showed good promise for the task of corallite center detection, but through limitations inherent in YOLO v3 the region shapes and orientations can only be approximated, and thus, effective translation to a three-dimensional representation is limited. With this in mind, it was decided that a segmentation approach would be better suited to produce a prediction output which could more effectively be used to describe the topology of the corallite regions.

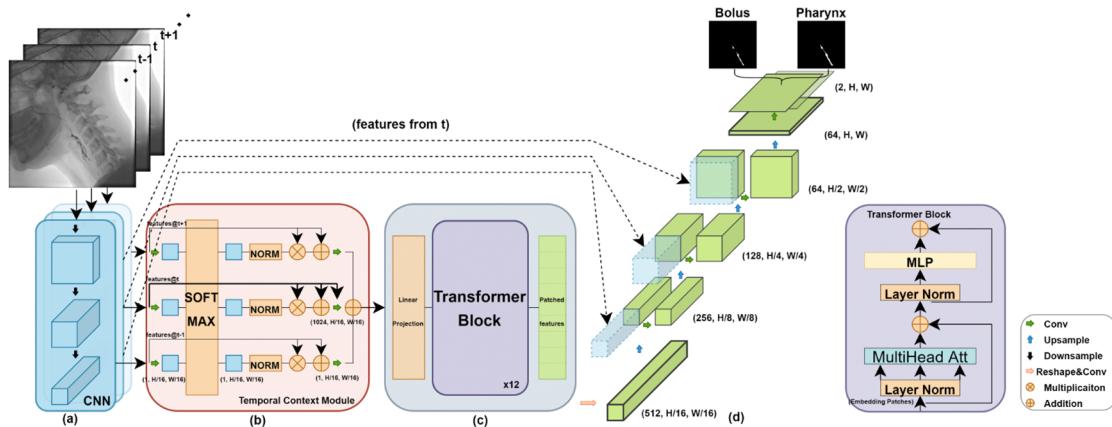
Trans-UNet: Developed in 2018 for medical image segmentation, Trans-UNet [5] was one of the first vision models in the medical domain to combine ViT’s in the UNet framework. Motivated by the loss of spatial feature resolution inherent in Transformers, the combined approach seeks to take ‘the best of both worlds’ by up-sampling the self-attentive features encoded by the Transformer in combination with the spatially context aware encoding of the CNN, via skip connections.

The implementation uses a combination of ResNet-50 [15] and ViT [10], both pre-trained on ImageNet [8]. Ablation studies showed best results using 3 skip connections with an input size of 512×512 and patch size of 8. However, it is noted that the computational cost is much higher than with inputs of 224×224 and patches of 16, for little extra reward, and thus the lower resolution inputs serve as preferable parameters.

Video Trans-UNet: Work done in [42] expands on the success of [5] by introducing a Temporal Context Module (TCM) which provides feature blending across multiple image frames, in this case, video intended for medical diagnosis.

Much of the architecture is consistent with [5], however the input is in the form $\mathbf{x} \in \mathcal{R}^{T \times H \times W}$, where T is time, ie. the number of sequential video frames to be used for temporal encoding. The TCM introduced in Video Trans-UNet is a trainable self-attention module which blends a linear projection of the ResNet-50 encoding across the temporal domain $t \in T$, thus encoding temporal information into a singular representation to serve as input to the Transformer part of the architecture. Ablation studies showed that an optimal snippet length for this application was 5, and naturally, larger snippets resulted in larger computational requirements.

Fig. 2.12 shows the architecture diagram presented in the paper for quick reference. This project will leverage the success of this work by adapting this pipeline for single-class corallite region segmentation, considering sequential μ CT slices as the ‘temporal’ dimension, for encoding via the TCM. A full description of the final model architecture used in this thesis is presented in 3.2.1.



(a) Multi-frame ResNet-50-based feature extractor; (b) Temporal Context Module for temporal feature blending across frames; (c) Vision Transformer Block for non-local attention-based learning of multi-frame encoded input; (d) Cascaded expansive decoder with skip connections as used in original UNet architectures, however, here with multiple prediction heads co-learning the two instances of clinical interest.

Figure 2.12: **Video Trans-UNet Architecture:** The architecture diagram of the Video Trans-UNet pipeline, as presented in [42].

Chapter 3

Project Implementation and Execution

3.1 Novel Volumetric Data

The first section of this chapter will focus on the dataset used for this project. This comprises data provided courtesy of the 4D-Reef Project¹, The Natural History Museum and the University of Bristol School of Earth Sciences. In addition, this chapter introduces a novel set of fully annotated slices which captures the full extent of omni-directional corallite growth.

The first section will provide a set of tables summarising the datasets used by this thesis, and some key properties of note. The subsequent discussion will cover the processes used to reshape the data, as volumetric inputs, for the deep-learning pipeline, and augmentations used at training time.

3.1.1 Dataset Summary

The following tables serve to summarise, in detail, the full extent of the datasets used throughout this dissertation as a means for quick reference where required. Table 3.1 provides a summary of the raw data, by colony, and scan projection axis. Table 3.2 provides a detailed summary of the novel annotations provided by this project, in the form of strongly labelled slices. Finally, Table 3.3 summarises the separation of datasets with respect to training and testing.

Coral Dataset	Total Slices	Annotated Slices	Annotation Quality	Notes
Porites Naturalis '6785' Unknown projection axis	60	33	Weak	Non-contiguous scans Incomplete colony
Porites Naturalis '6781' Growth projection axis	248	2	Strong	Full colony
Porites Naturalis '6781' Orthographic projection axis	388	1	Strong	Full colony
Porites Astraeiformis Unknown projection axis	1	1	Strong	Different species Unknown μ CT parameters Single slice dataset

Table 3.1: **Raw Data Summary:** Summary table of the raw data used for this project provided courtesy of the 4D-Reef Project, The Natural History Museum and the University of Bristol School of Earth Sciences. The strength of the annotation quality refers to completeness of labelling, as discussed in section 3.1.2. Dataset colour coding for ease of reference to the annotation details set out in Table 3.2.

¹<https://www.4d-reef.eu/>

3.1. NOVEL VOLUMETRIC DATA

Slice Details	Resolution (px)	Total Corallites	Total Labelled Pixels	Circular Regions	Elliptical Regions
Porites 6781 Growth Slice #1279	4023×3487	1847	563079	1271	630
Porites 6781 Growth Slice #2499	4023×3487	1684	378593	1329	292
Porites 6781 Ortho Slice #1003	3487×3432	1043	354757	601	442
Porites Astraeiformis	2809×2149	3838	989105	2434	1404

Table 3.2: **Annotation Details:** Detailed breakdown of the properties of each of the novel annotated slices provided by this project. The circular and elliptical regions are approximate counts. A region is considered circular if the absolute difference between the lengths of the major and minor axes is less than 10 pixels. An illustration of these axes are presented in Fig 3.27. Colour coding indicates the colony dataset from which each slice was drawn, according to descriptions provided in Table 3.1.

Model	Training Data	Testing Data
Naive	Naturalis 6785 (80%)	Naturalis 6785 (20%) (same training/testing domain)
Fine-Tuned Models	Pre-train: Nat. 6785 (100%) FT: Nat. 6781 Growth #1279 (100%)	(1): Nat. 6781 Growth #2499 (100%) (2): Nat. 6781 Ortho #1003 (100%) (3): Astraeiformis (100%) (complete separation between training/testing domains)

Table 3.3: **Data Separation:** Separation of datasets for testing and training for both the naive model, and for later, fine-tuned models. The naive model is trained and tested on the same dataset, the weakly annotated Naturalis (Nat.) 6785, separated by an 80/20 split with no overlap. The fine-tuned models are pre-trained on 100% of the Naturalis 6785 data, and then fine-tuned on one of the complete Naturalis 6781 slices: #1279 from the Growth axis projection. Testing sets for the fine tuned models include: (1): Naturalis slice #2499, also from the Growth projection axis, but well separated by distance within the scan volume. (2): Slice #1003 from the Ortho projection axis, unknown to the trained models. (3): A slice from the Astraeiformis species of the Porites genus, unknown to the trained models, with completely unknown scan parameters, position in the colony, and axis of projection.

3.1.2 Dataset

This project was provided with a collection of μ CT scans from multiple colonies of the Porites genus. Each colony within the collection has typically been scanned from two principle axis labelled 'Growth' and 'Orthographic'. Each scan axis comprises a collection of cross section slices. Each slice is formed by the aggregation of five sequential μ CT layers, to provide better visual corallite region definition.

Annotations: The University of Bristol School of Earth Sciences provided annotations for a subset of slices from one of the colonies (Naturalis 6785), an example of which is presented in Fig. 3.2. These annotations will form the foundation of the data set used for this project, but do present some limitations of note:

- All annotated data comes from the same colony and may thus limit the models ability to predict on unseen colonies.
- Annotations are provided for only the circular structural regions of the image, where the corallite growth aligns with the projection axis of the μ CT scan, ie. grows vertically with respect to the image origin. Corallite growth direction is influenced by environmental factors, and is therefore non-linear [25]. The longer elliptical regions represent growth which is perpendicular to the projection axis, as illustrated by Fig. 3.1. This growth information is not captured by the annotation set, and therefore limits the possibilities of deep-learning predictions for omni-directional corallite growth.

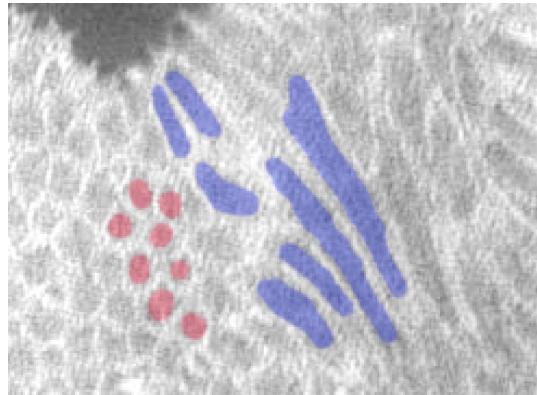


Figure 3.1: Corallite Region Variations: An illustrative example of the differences in growth direction within a μ CT slice. Circular structures (red) represent growth aligned to the projection axis, elliptical structures (blue) represent growth perpendicular to the projection axis

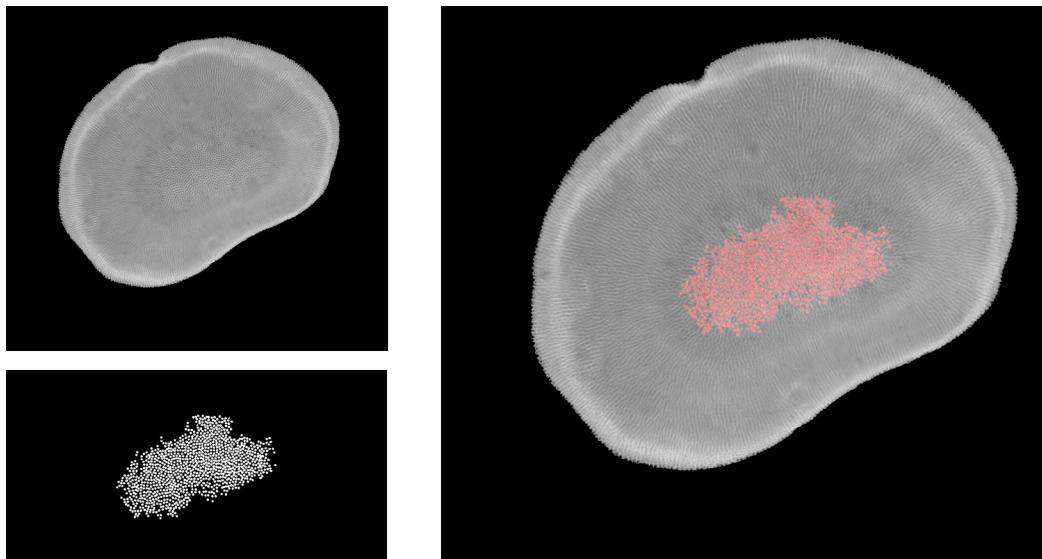


Figure 3.2: Inherited Annotation Set: An example of the annotations inherited by this project. Top-Left: Raw μ CT data, Bottom Left: Annotations, Right: Composite where annotations are presented in red. Note that the annotations only cover a subset of the total number of corallite regions, specifically, only those with circular geometric shape.

Extended Annotation Set: In light of these limitations, further annotations were made using Adobe Photoshop² for four complete slices, from two colonies of the *Porites* genus, presented in Fig. 3.3. Three slices from the *Porites Naturalis* '6781' colony: Two slices from the μ CT projection axis labelled 'Growth', and another from the projection axis labelled 'Ortho'. One slice from a *Porites Astraeiformis* colony, from unknown projection direction, intended to test for inter-species generalisation performance.

It is important to note that each complete colony scan was produced using different μ CT equipment and parameters, leading to significant variance in the quality, resolution, contrast and luminosity of the scan sets. For the purpose of this project, these additional annotations will provide the data to explore the use of transfer-learning techniques to improve omni-directional corallite growth predictions, and generalisation across colonies of the *Porites* genus.

Summary: This project presents a novel set of 8,412 corallite region annotations across two new species of the *Porites* genus. This data will be used for model fine-tuning, and inference testing on unseen axis from known colonies, and on entirely unseen colonies captured with unknown μ CT equipment parameters. It also serves to bolster the amount of gold-standard annotations for this domain, providing full-coverage labelling of the geometric structure of corallite regions in each slice.

²<https://www.adobe.com/uk/products/photoshop.html>

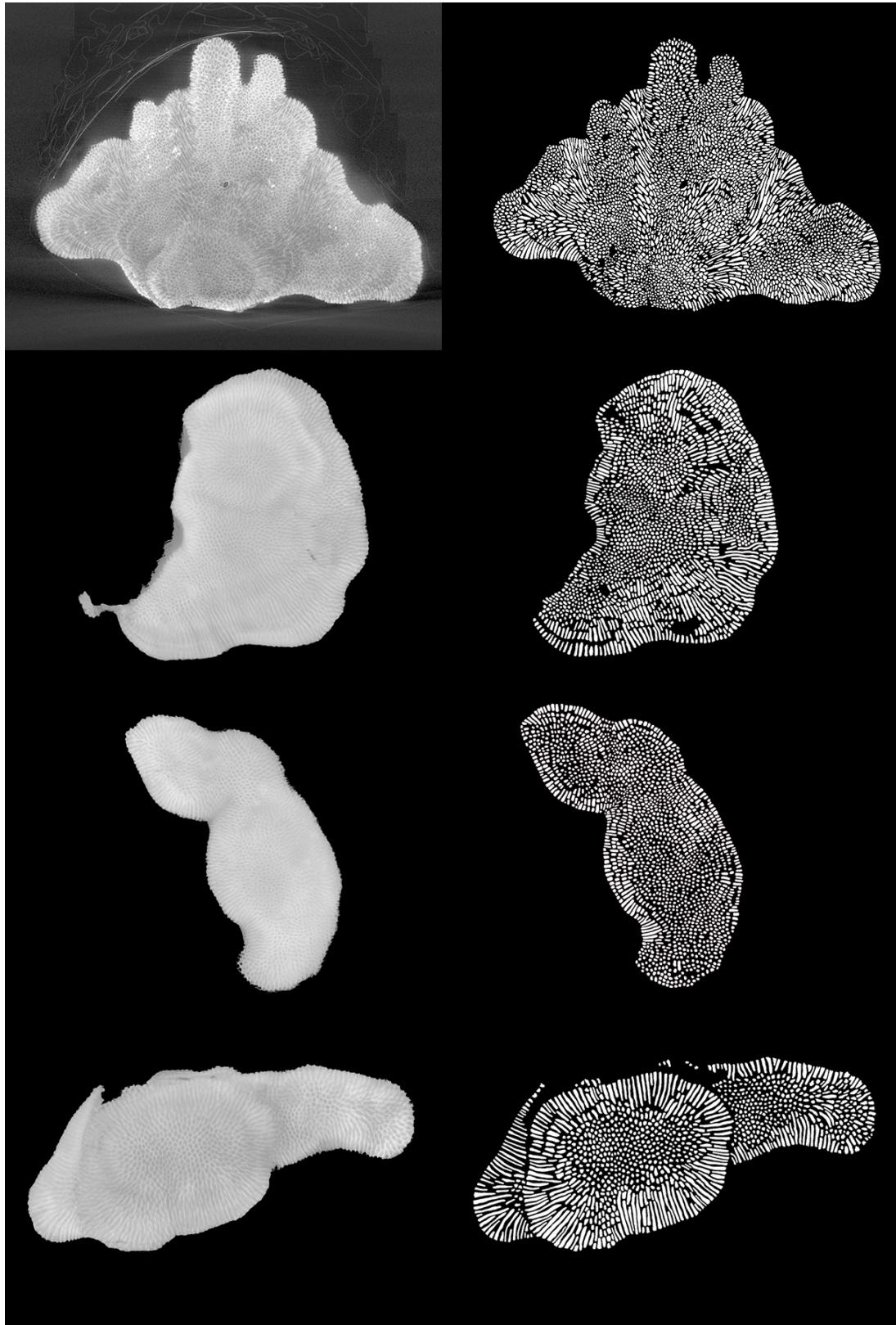


Figure 3.3: **Novel Annotation Set:** Additional annotations produced for this project. Left: μ CT scan, Right: Novel Annotation. Top to Bottom: *Porites Astraeiformis*, *Porites Naturalis* 6781 Growth slice#1279, *Porites Naturalis* 6781 Growth slice#2499, *Porites Naturalis* 6781 Ortho slice#1003

3.1.3 Data Preparation

Across all datasets, each μ CT image slice is invariably of high resolution in order to capture as much of the colony detail as possible (eg. 4023×3478 pixels). A common bottleneck when dealing with image-based deep learning tasks is limitations on available GPU memory, an issue which becomes more pronounced as the size of the input grows [27]. One technique to manage this is to provide relatively small input sizes to the deep learning framework. Following findings in [5], this project will be using inputs of size $W \times H = 224 \times 224$. For many of the images in the dataset this is a reduction of approximately $400 \times$ resulting in significant loss of information, as shown in Fig. 3.4.

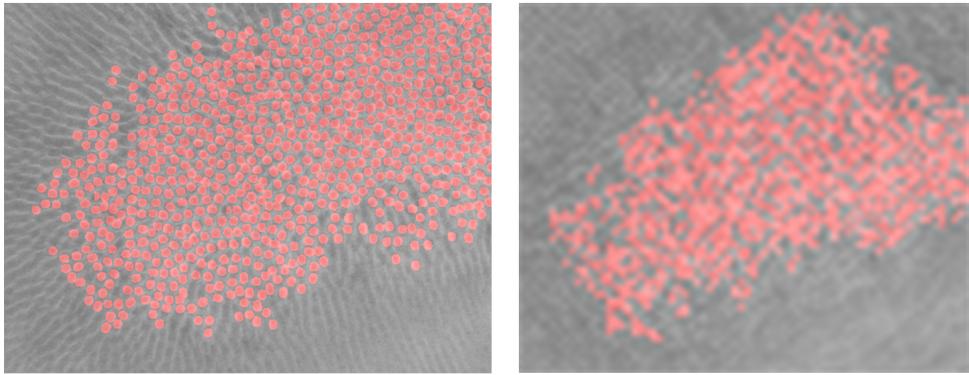


Figure 3.4: **Reduction Issues:** An illustrative example of the loss of data resulting from input size reduction for deep-learning. Left: A section of full resolution (2210×1990) Porites Naturalis 6785 annotation, Right: The same annotation after reduction to 224×224 .

Tiling: Corallites are very small, even at 4023×3487 resolution, a corallite region is approximately 15×15 pixels. Given the size, and the large number of corallite regions within a single slice, the relationships tend to be localised to a small area. With this in mind, a decision was made to sacrifice some of the wider contextual information in order to retain more fine-grained detail. This is achieved by tiling each slice using a sliding window of size $W \times H$ with step size k . An illustrative example of this tiling process is presented in Fig. 3.5, and example tiled image/annotation composites are presented in Fig. 3.6. In addition to maintaining image and annotation resolution, a tiling approach increases the volume of annotated data for training. This is subject to risk of over-fitting if there is overlap in the sliding window, ie. $k < W$ and/or $k < H$, therefore care must be taken to ensure that there is proper separation between training and testing datasets.

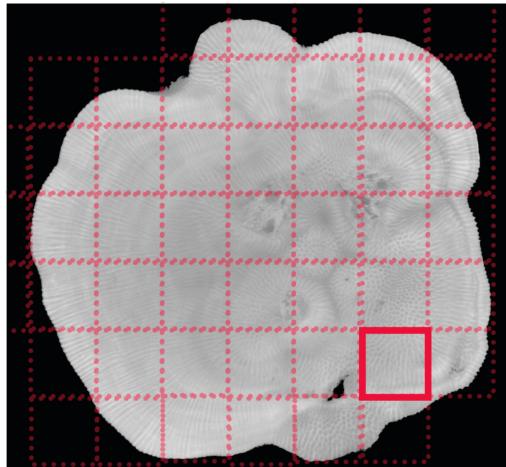


Figure 3.5: **Tiling Approach:** An illustrative example of a μ CT slice tiled with window $k \times k$ and step size k .

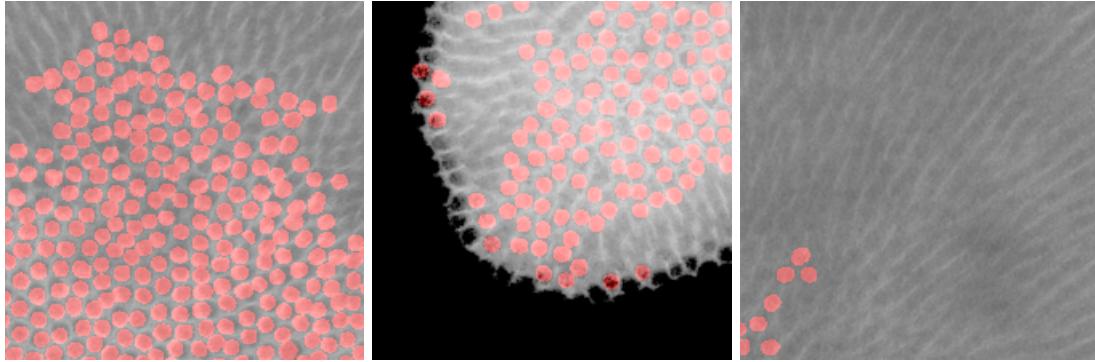


Figure 3.6: **Composite Tile Examples:** Example 224×224 tiled image and annotation composites from the Porites Naturalis 6785 dataset.

Volumetric Snippets: In order to leverage the volumetric information in the data and provide data to the TCM of the architecture (section 3.2), volumetric inputs are constructed as snippets $D \times W \times H$ consisting of sequential μ CT slices of depth D , where D has a parity of 1 and the center image ($\lfloor \frac{D}{2} \rfloor + 1$) contains the annotation, as illustrated in Fig. 3.7. A baseline snippet depth of five was selected for this project based on the findings in [42].

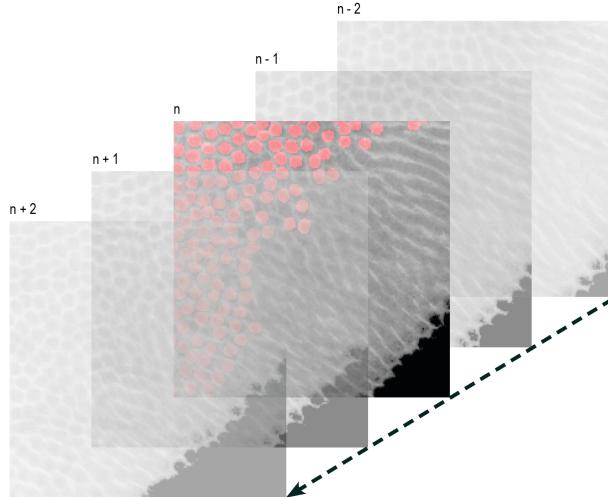


Figure 3.7: **Volumetric Snippets:** An example snippet of dimensions $5 \times 224 \times 224$. Note that the annotation for slice n is in the center of the volume.

3.1.4 Data Separation

Machine learning models rely on separating the data into three key groups: training, validation and testing, for example, in proportions of 70%, 20%, 10% respectively [3]. The training data is used as input to the model, and serves as a target used in calculating the loss for back propagation. The validation data is used at the end of each training cycle to determine how well the model performs on a set of data that has not influenced the weight updates.

Performance on the validation set is typically an indicator for when the optimisation has converged at a minima. The testing data is used outside of the training exercise, once the model has converged, to analyse how well the model performs on completely unseen data, indicating its ability to generalise within the domain of that test set.

One of the key goals underpinning this project is to produce a model which can predict corallite regions across different colonies of the Porites genus. As discussed in section 2.2.6, a two stage fine-tuning process will be conducted to generate a baseline model using the weakly annotated Porites 6785 data, with further fine-tuning on the strongly annotated Porites 6781 data. Validation sets used for model training are generated by reserving 10% of randomly selected training data when initialising the model at train time. The data separation breakdowns are presented in Table 3.3.

3.1.5 Augmentation

Many deep-learning models suffer from risks of over-fitting, where the model learns to predict the training and validation sets so precisely, that it under performs when predicting on unseen, real-world data. A frequently used practise to managing over-fitting is augmentation of the training data. This augmentation can take many forms, and research is ongoing to further understand and develop new augmentation techniques which boost a models ability to generalise. It is important that augmentations are only applied to training data, and that the validation set remains in its true form, as this is used to determine convergence in the target domain.

For the purpose of this project, three simple transformations will be used, selected at random, during training, for each input snippet:

- Apply a random rotation to the snippet between $[-20, 20]$ degrees.
- Apply a random rotation and flip to the image by rotating the snippet by one of $\{0, 90, 180, 270\}$ degrees and then flipping across a randomly selected axis $\{0, 1\}$.
- Add Gaussian noise to the image with parameters: $\mu = 0$ and random variance in range $[0, 0.5]$, introduced in later models.

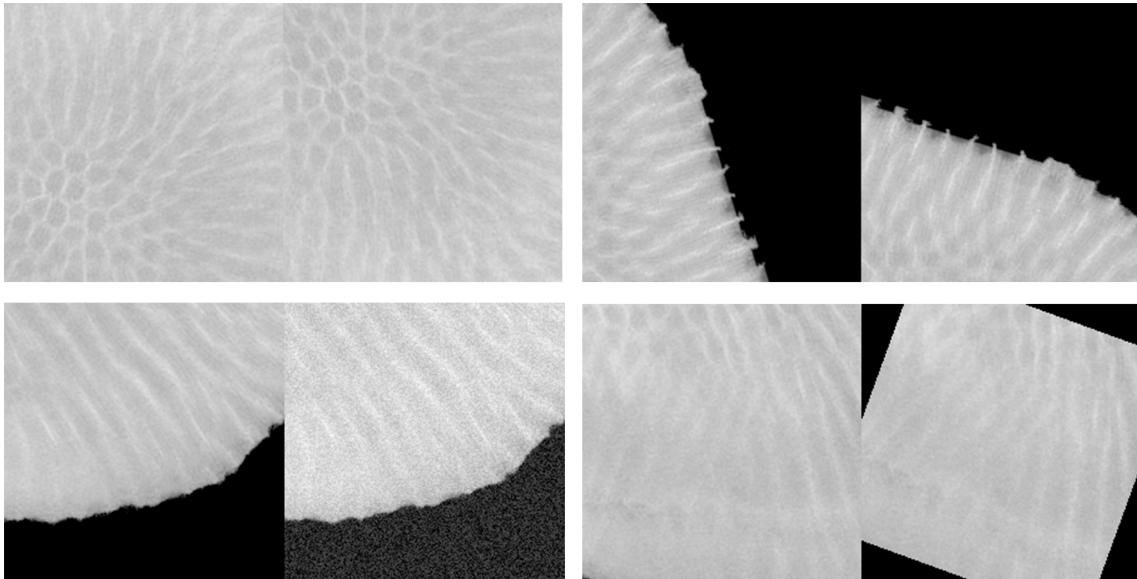


Figure 3.8: **Training Augmentations:** Some example visualisations of the augmentations applied to the training data. For each image pair the original image is presented left with the resulting transformation presented right. Top Left: Rotation of 180 degrees. Top Right: Rotation of 90 degrees and flip. Bottom Left: Addition of Gaussian noise. Bottom Right: Rotation between $[-20, 20]$ degrees.

3.1.6 Summary

This section covered the various stages of data preparation required to create a viable input set suitable for volumetric deep-learning prediction. The importance of annotations which described the full range of corallite regions were discussed, motivating the introduction of a novel dataset of over 8,400 labelled corallites. Input tiling techniques were described to help resolve issues of information loss, and a full description of the formation of volumetric snippets was provided. A description of the data augmentation techniques used on the training data was also provided. Discussions in section 3.2, and corresponding evaluations in section 4.2 will explore the influence of the novel annotations, augmentations to training data, and tiling step-size on prediction performance.

3.2 Deep Learning Pipeline

This section details the investigations carried out in adapting and optimising the original pipeline used in [42], for the task of volumetric corallite segmentation. Some adjustments to the model architecture precede these discussion points. These were relatively minor changes made to ensure that the inputs and outputs conformed to the requirements of the task. Specifically, the model was adapted from a multi-class segmentation pipeline, to a single-class segmentation pipeline.

First, a naive model is introduced using the weakly annotated Naturalis 6785 dataset to evaluate a 'first-touch' baseline. Next, a variety of fine-tuning implementations are discussed, including the introduction of a state-of-the-art baseline for this project, used as a foundation for all future model evaluations. Subsequently, the implementation of the custom topological loss function is discussed with particular regard to the techniques used to include it as complement to the existing loss. Lastly, further experiments to improve the models ability to generalise are explored. Summary results are discussed in brief, however detailed evaluations of performance are covered in Chapter 4.

3.2.1 Corallite Prediction Architecture

The work in this thesis, in part, seeks to answer the question (DL1 1.1) on whether the temporally aware Video Trans-UNet architecture defined in [42] is transferable to the domain of corallite region segmentation. The underpinning data is similar for this thesis (micro-computed tomography imaging) and the fundamental task of segmentation is aligned. The expectation is that with some adjustments to the overall loss function, and application of appropriate transfer-learning techniques, this framework should serve as a good basis for volumetric corallite region prediction.

Architecture: Following the work of both [5] and [42] this project uses a ResNet-50 + ViT backbone, pre-trained on ImageNet [8], with input size 224×224 , patch size of 16, and 3x skip connections. The model accepts inputs in form $\mathbf{x} \in \mathcal{R}^{D \times H \times W}$, where $H \times W$ defines the pixel resolution and D is the number of sequential μ CT slices. Features are blended across D , using the TCM described in [42], with the resulting encoding passed as input to the 12-layer ViT. Up-sampling follows [5], in a cascaded approach, incorporating the skipped ResNet-50 encodings. A full architecture diagram is presented in Fig. 3.9.

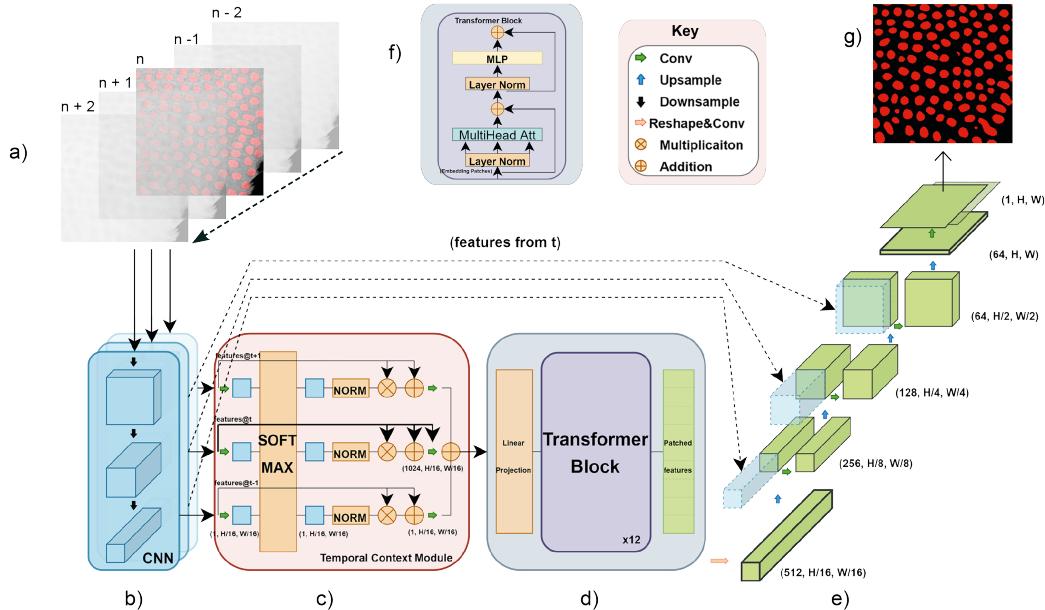


Figure 3.9: **Full architecture diagram.** (a): A $5 \times 224 \times 224$ input volume is provided to (b): A pre-trained ResNet-50 CNN encoder with 3x skip connections. (c): Feature encodings are blended across the volumetric dimension before begin passed as input to (d): a pre-trained 12 layer ViT with Multi Head Attention and MLP transformer block components (f). (e): Finally, through cascaded upsampling, incorporating the skipped encodings a final segmentation output of size 224×224 is produced (g). Adapted from architecture flow diagram used in [42].

Hyper-parameters: In alignment with the findings from [42], the following hyper-parameters were used throughout all experiments, unless otherwise stated.

- Learning Rate: 0.0005 (adaptive via Adam optimiser)
- Loss: $\frac{1}{2}(\mathcal{L}_{BCE} + \mathcal{L}_{DSC})$
- Batch Size: 5

Naive Model: Using the architecture defined above, and the weakly annotated training and testing sets defined in section 3.1.4, a naive model was trained for 300 epochs, achieving validation convergence at ~ 150 epochs with a dice similarity accuracy of 0.59. Training loss and accuracy graphs are presented in Fig. 3.10 with results presented in Table 4.2. When predicting corallite regions on slices of the same colony this model achieves a mean DSC of 0.56. Some example outputs are presented in Fig. 3.11. Performance is particularly poor on unseen colonies, as seen in Fig. 3.12, even for regions with same geometric shapes as the weakly annotated training data, ie. circular. It is important to note the absence of predictions from sub-regions of each tile, a result of incomplete slice annotations from the Naturalis 6785 dataset. These results strongly highlight the limitations of the Naturalis 6785 data for both omnidirectional region prediction, and ability to generalise between colonies and species within the Porites genus.

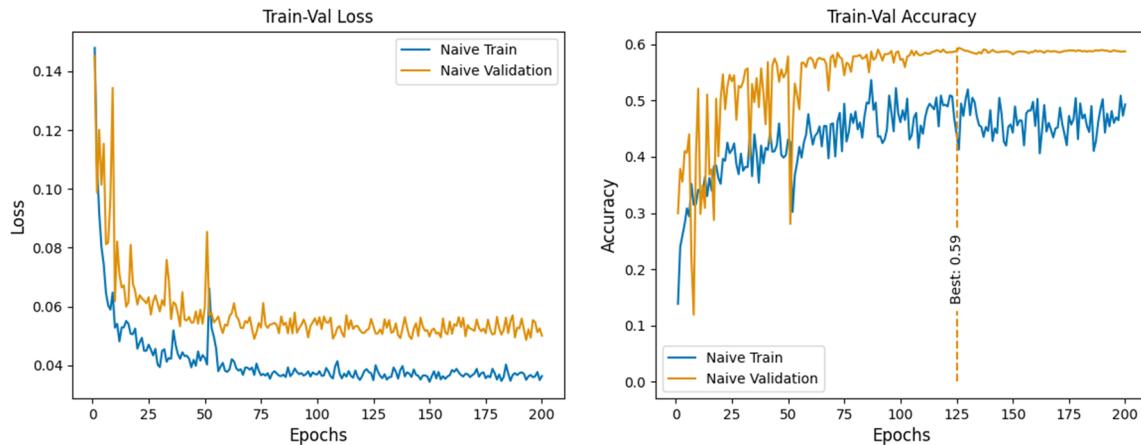


Figure 3.10: **Naive Model Training:** Loss and DSC accuracy graphs generated during training of the naive model. Model convergence is reached after ~ 150 epochs with a validation accuracy of 0.59. Note that despite training loss being less than validation, validation accuracy is higher.

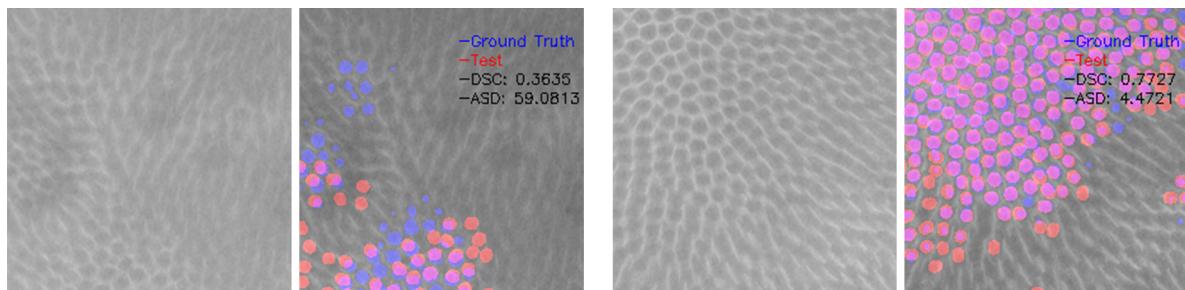


Figure 3.11: **Naive Model Inference - Porites Naturalis 6785:** Naive model predictions on slices from the same coral colony as the training data. For each pair the left image is the input tile, and the right a composite of the model prediction and ground-truth labels, where red is prediction segmentation and blue is ground-truth. **Left Pair:** Slice #395: Mean Dice 0.36. **Right Pair:** Slice #415 (bottom): Mean Dice 0.77.

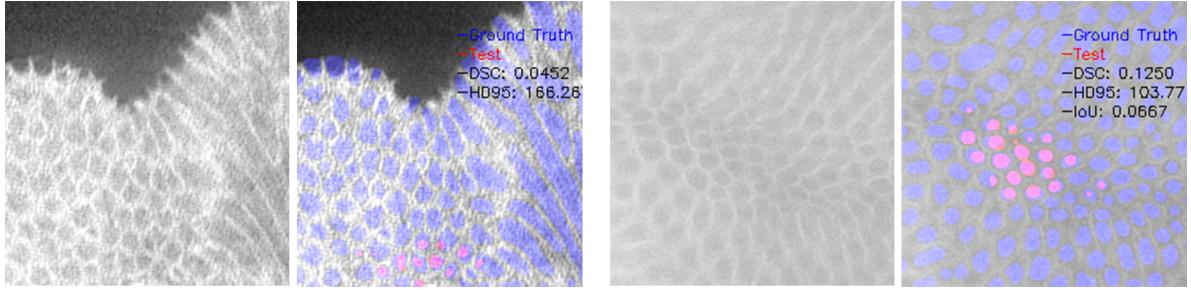


Figure 3.12: **Naive Model inference - Unknown Colonies:** Naive model predictions on slices from different coral colonies of the *Porites* genus. For each pair the left image is the input tile, and the right a composite of the model prediction and ground-truth labels, where red is prediction segmentation and blue is ground-truth. **Left Pair:** *Porites Astraeiformis*: Mean Dice 0.045. **Right Pair:** *Porites Naturalis* 6781 Growth slice #1279: Mean Dice 0.1250.

3.2.2 Fine-Tuning

In an effort to establish a state-of-the-art (SotA) corallite segmentation model, this thesis explores the impacts of transfer-learning techniques, first defining an SotA baseline by leveraging fine-tuning techniques to utilise the novel datasets introduced by this project. Experiments are subsequently conducted to evaluate the impacts of further expanding the volume of strongly annotated data by reducing the sliding window step-size when tiling the input during pre-processing (section 3.1.3).

Fine-Tuned Baseline: A SotA baseline model was pre-trained on 100% of the Naturalis 6785 weakly annotated data, and then fine-tuned on the fully annotated Naturalis 6781 Growth slice #01279 with freezing of the TCM. Training and validation curves plateau very quickly to a minimum loss of 0.09 and peak accuracy of 0.78. The loss curve, presented in Fig. 3.13, suggests a good model fit and that efforts to improve overall accuracy might be best spent on addressing shortcomings in the training data. Final testing performance results are presented in Table 4.2. The baseline model has a mean Dice similarity of 0.76 on an unseen slice from the same colony and same projection axis, with diminishing returns as the prediction domain moves further away from the training data; Dice similarities of 0.67 for slice #1003 from the same colony but different projection axis, and 0.61 for the *Astraeiformis* slice, which is of a different *Porites* species to that of the training data. Example prediction vs. ground-truth visualisations for *Porites* 6781 are presented in Fig. 3.14 and Fig. 3.15.

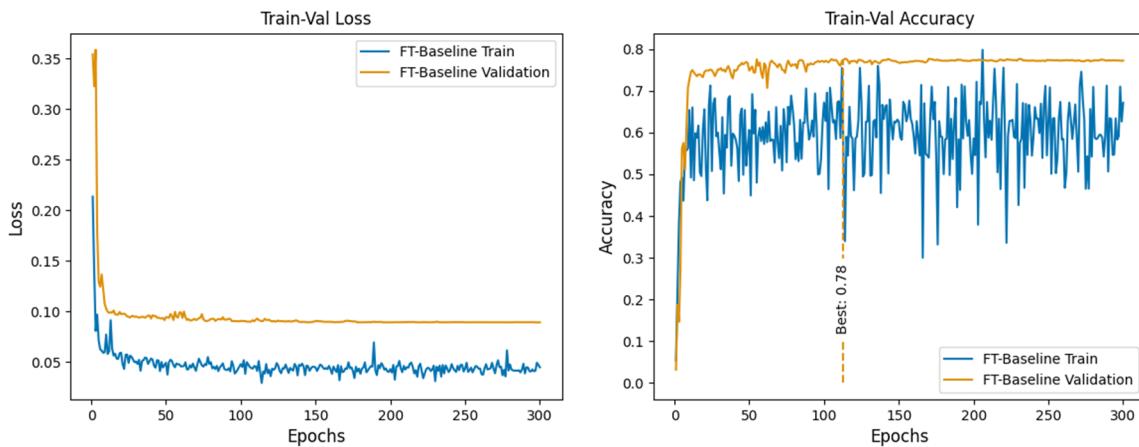


Figure 3.13: **Baseline Training:** Loss and Dice-similarity accuracy graphs generated during training of the baseline fine tuned model. Note that convergence occurs very rapidly, achieving a peak validation accuracy of 0.78.

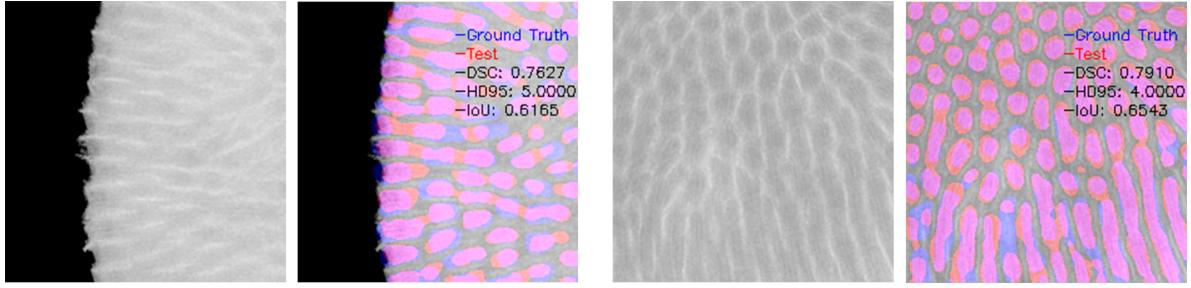


Figure 3.14: Baseline Inference - Known Colony: Prediction examples from two different test sets of the Naturalis 6781 colony using the baseline fine-tuned model. For each pair the left image is the input tile, and the right a composite of the model prediction and ground-truth labels, where red is prediction segmentation and blue is ground-truth. **Left Pair:** Growth slice #2499, which is the same projection axis as the training set. **Right Pair:** Ortho slice #1003, an unseen projection axis of the same colony.

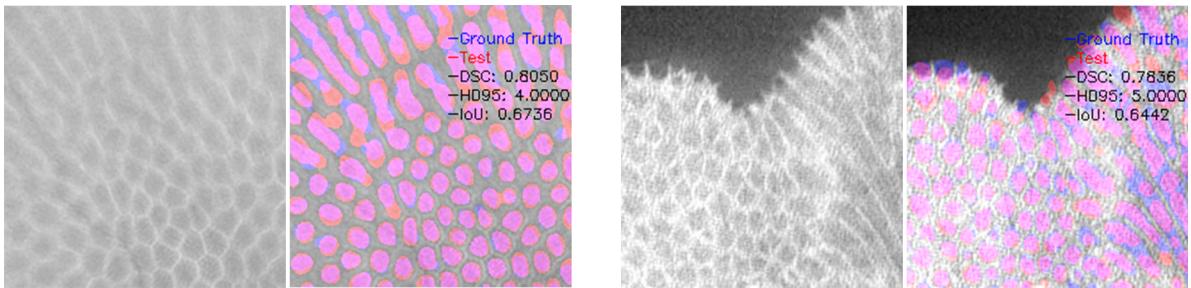


Figure 3.15: Baseline Inference - Unseen Axis/Colony: Prediction examples from the Ortho and Astraeiformis test sets 3.1.1. For each pair the left image is the input tile, and the right a composite of the model prediction and ground-truth labels, where red is prediction segmentation and blue is ground-truth. **Left Pair:** Ortho slice #1003, which is the same colony as the training data, but from a different projection axis. **Right Pair:** Astraeiformis, a completely different species to the training data. Note that the prediction coverage is dramatically improved in comparison to the naive implementation (Fig. 3.12).

Expanded Training Data: In order to expand the available training data to aid driving model performance upwards, two new sets were generated by reducing the step size of the sliding window when tiling the raw images prior to snippet construction, as described in section 3.1.3. The first set, labelled 'Lrg', uses a sliding window step size of $k = 112$, giving a training set size of 3x the baseline set. The second set, labelled 'VLrg', used a step size of $k = 50$, giving a training set size of 16x the baseline set. These additional training examples were subject to the same augmentation treatment as described in section 3.1.5.

The impacts of this expansion on the training accuracy are pronounced, taking the baseline of 0.77 to 0.93 and 0.94 for the Lrg and VLrg sets, respectively. Dice similarities for the test data are marginally improved by 1% on the known axis (Growth slice #2499) for both models, and improvements extend further when testing on the unseen axis (slice #1003): 2% and 4% for the Lrg and VLrg models.

The effects of over-fitting, due to the overlap in the training data, is evident when testing on the Astraeiformis species, where reductions of up to -5% in Dice similarity against the baseline model are observed. This results in numerous false-positives, as shown in Fig. 3.18.

Training curves for both of these models are presented as a composite for brevity in Fig. 3.16, with performance results presented in Table 4.2. It is worth noting that these models begin to identify some of the weaknesses of DSC as a metric for this task, as it does not describe well the region separation accuracy between corallites. This is illustrated in Fig. 3.17, where an area has been selected to illustrate the improvements to region separation.

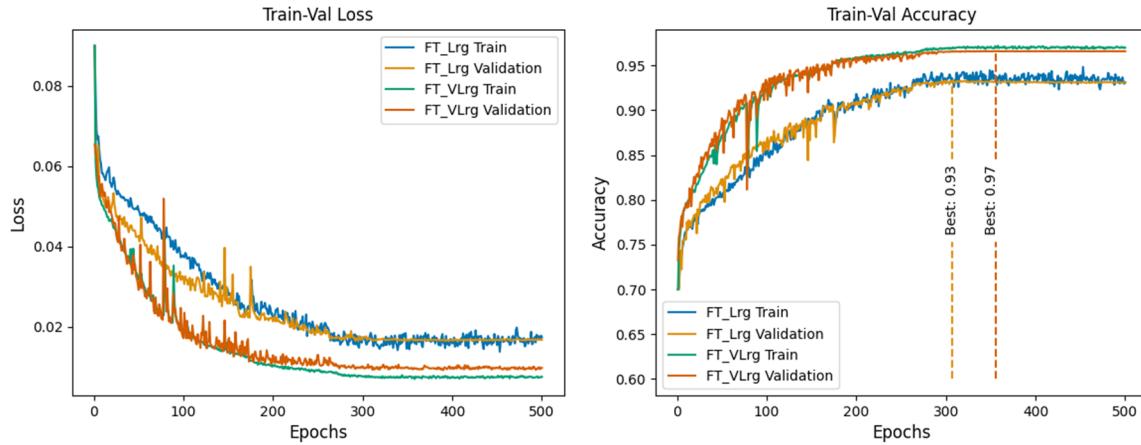


Figure 3.16: **Fine-Tuning Training Performance:** Composite Loss and DSC accuracy graphs generated during training of fine-tuned models using the Lrg and VLrg training sets respectively. Note that training and validation accuracy's for both fine-tuned models have almost converged. Peak DSC accuracy is noted for both models; Lrg: 0.93, VLrg: 0.97.

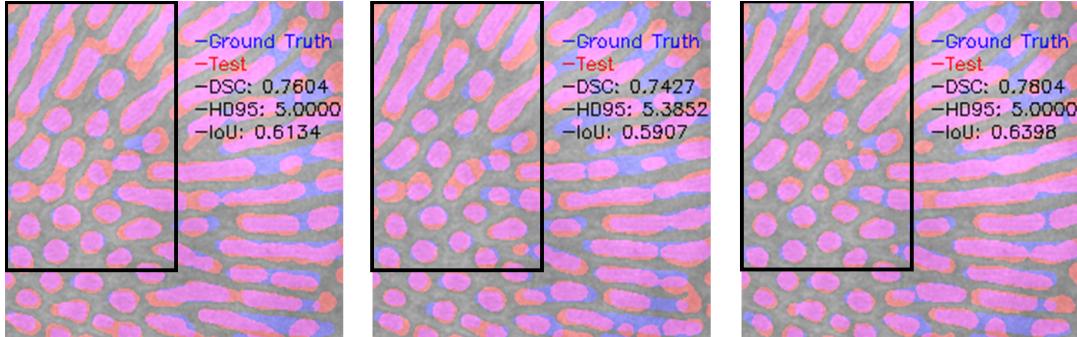


Figure 3.17: **Fine-Tuning Inference Comparison - Unseen Axis:** Comparative composite ground-truth (blue) overlaid with prediction (red) from the Ortho testing set 3.1.1. From left to right: Baseline, FT-Lrg, FT-VLrg. Note that the VLrg model begins to show better performance at predicting the longer corallite regions which run perpendicular to the projection axis. The black box regions highlight an area in which the corallite region separation improves between models, but which is not necessarily reflected in the DSC similarity scores.

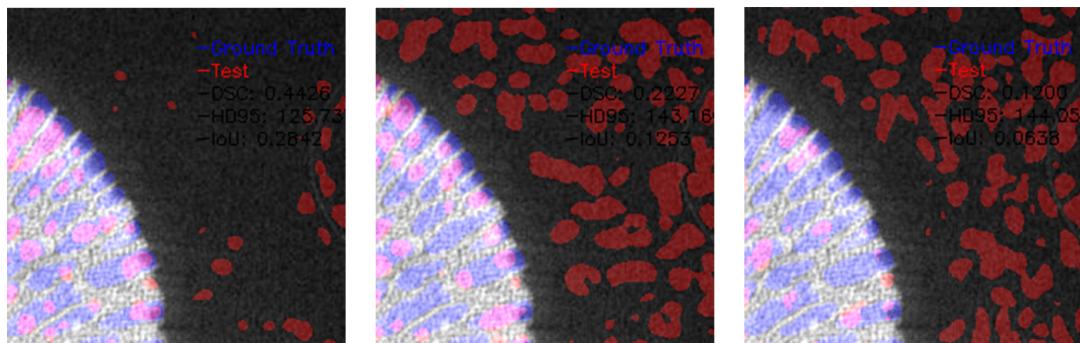


Figure 3.18: **Fine-Tuning Inference Comparison - Unseen Species:** Comparative composite ground-truth (blue) overlaid with prediction (red) from the Astraeiformis testing set 3.1.1. From left to right: Baseline, FT-Lrg, FT-VLrg. Note that the larger models are displaying significant issues associated with over-fitting, resulting in many more false-positives where there is background noise in the scan data.

3.2.3 Topological Loss Implementation

One of the key goals of this project is to generate a three-dimensional model of corallite growth within the colony. In order to do this with reasonable accuracy, it is important that the model predicts corallite regions with the correct topology, to ensure that the resulting model geometry is an accurate reflection of the true colony structure. Whilst fine-tuning produced good results in terms of image similarity, inspection of the prediction outputs showed some failures in region separation, resulting in topological inaccuracies. An example of this is illustrated in Fig. 3.19. Of principle note is the inter-connected region predictions, where in the right image, a red segment (false positive) connects two purple regions (true positive). In the loss map on the left, these areas are indicated by the brighter regions, representing a larger error. An investigation was carried out to understand if training the model for topological correctness, in addition to the Dice and BCE similarity measures, would have a positive impact on prediction accuracy.

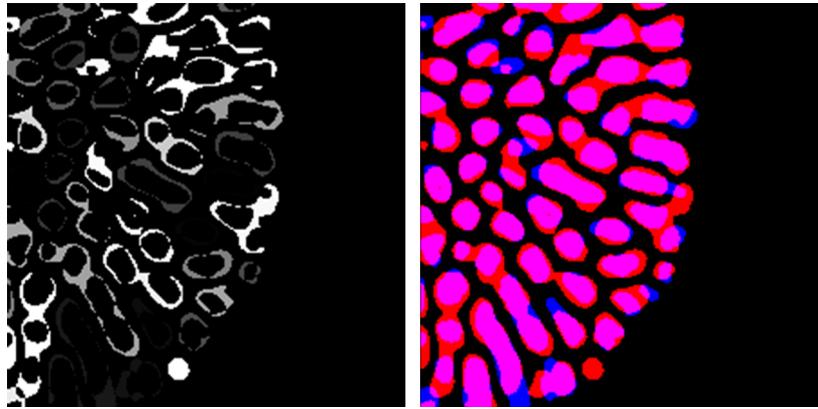


Figure 3.19: **Topological Loss Visualisation:** An example of topological error in model predictions, similar to Fig. 2.8, where model segmentations are expressed as connected regions, instead of individual corallites. Topological error map is presented on the left, composite ground-truth (blue) and predictions (red) on the right. The topological loss is designed to proportionally discourage these kinds of prediction errors.

Error Function: Using the Scikit-Image module *measure*³, the sets of all contiguous prediction regions P and ground-truth regions L are used to generate a pixel-wise error map, based on the differences in the area of each prediction region and its nearest neighbour ground-truth region. Formally, given prediction region p with area A_p , and nearest-neighbour label region l with area A_l :

$$p_\epsilon = \min \left[\left| 1 - e^{\frac{A_p - A_l}{A_p}} \right|, 1 \right]$$

Thus, each pixel belonging to region p is assigned the region error p_ϵ , if it is not a member of the ground truth label l .

Note that this formulation ensures that the errors tend to 0 when the region areas are similar, and that the rate at which the error penalty increases for larger false positives is higher than for those which are smaller. A graphical illustration of this is presented in Fig. 3.20. The algorithm for producing the error map is presented in Algorithm 3.1, and some example error map with associated prediction/ground-truth overlay are presented in Fig. 2.8 and Fig. 3.19.

The final loss calculation is the mean element-wise (pixel-by-pixel) multiplication between the error map \mathcal{E} and the model segmentation predictions S , offset by +1. Normalisation is applied by an adjusted logistic sigmoid function $2(\sigma(x) - 0.5)$:

$$\mathcal{L}_{TOPO} = \frac{1}{N} \sum_n^N 2(\sigma(\mathcal{E}_n(S_n + 1)) - 0.5)$$

³<https://scikit-image.org/docs/stable/api/skimage.measure.html>

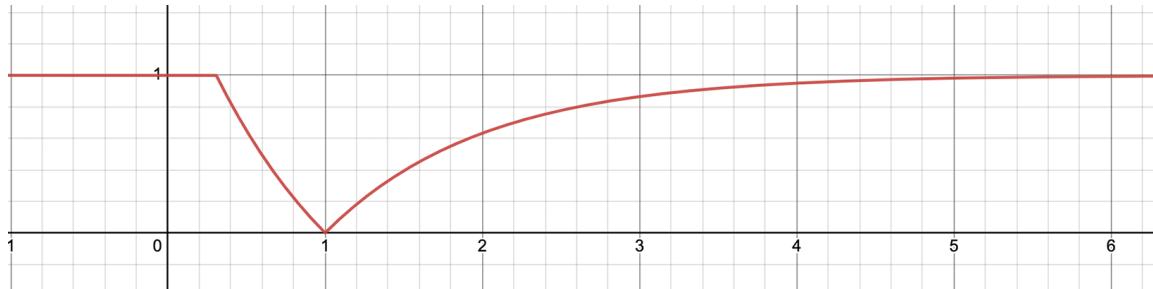


Figure 3.20: **Topological Error:** A graphical illustration of the error function $y = \min \left[\left| 1 - e^{\frac{A_p - A_l}{A_p}} \right|, 1 \right]$ where $x = \frac{A_l}{A_p}$. Note that the rate of change of the error is much more aggressive when $A_p > A_l$, as this is a principle marker of false positive segmentations, which typically manifest as interconnected regions.

Algorithm 3.1: Topological error calculation

```

CalculateTopologicalError(Prediction Regions: P, Label Regions: L, Error Map: E)
for pred ∈ P do
    error ← 0
    minDist ← ∞
    nn ← None
    pc ← pred.center
    for label ∈ L do
        lc ← label.center
        dist ←  $\sqrt{(pc_x - lc_x)^2 + (pc_y - lc_y)^2}$ 
        if dist < minDist then
            | minDist ← dist
            | nn ← label
        end
    end
    if  $\forall p \in pred.pixels : p \notin label.pixels$  then
        for  $\forall p \in pred.pixels$  do
            | E[p] ← 1
        end
    else
        Al ← label.area
        Ap ← prediction.area
        error ←  $\min \left[ \left| 1 - e^{\frac{A_p - A_l}{A_p}} \right|, 1 \right]$ 
        for  $\forall p \in pred.pixels$  do
            | if p  $\notin$  label.pixels and E[p] < error then
            | | E[p] ← error
            | end
        end
    end
end
return E

```

Integration to the Pipeline: The training pipeline was updated to include the new topological loss in addition to the existing BCE + DSC loss, as complement, after a warm-up period was completed. Specifically, when the number of training epochs was greater than n , the loss function was updated to:

$$\mathcal{L} = \frac{1}{2}(\mathcal{L}_{BCE} + \mathcal{L}_{DSC}) + \lambda \mathcal{L}_{TOPO}$$

Where λ is some scaling factor on the topological loss. Experiments were conducted to find a suitable λ , using a warm-up period of 50 epochs, chosen as it marked the point at which the validation loss was approximately half-way to convergence for the VLrg fine-tuned model discussed above.

Performance: Training performance of 0.97 DSC accuracy was maintained when $\lambda = 1$, but saw diminishing returns for other scalar values. Loss and DSC accuracy graphs for three of the experimental values are presented in Fig. 3.21. Regardless of a lack of improvement in validation accuracy's, significant improvement was observed in testing performance for unseen slices from a different projection axis to the training data, increasing DSC by a further +5%.

Importantly, the addition of topological loss shows improvements to the consistency and reliability of the topology of corallite region predictions. This will help to define regions more explicitly, providing stronger geometry within the prediction outputs, and thus aid the 3D model generation process. Some examples of these improvements are presented in Fig. 3.22.

It is worth noting that DSC scores for testing on the Astraeiformis data dropped by almost 0.1, indicating a significant decrease in inter-species generalisation accuracy. Detailed results for all testing sets are presented in Table 4.2.

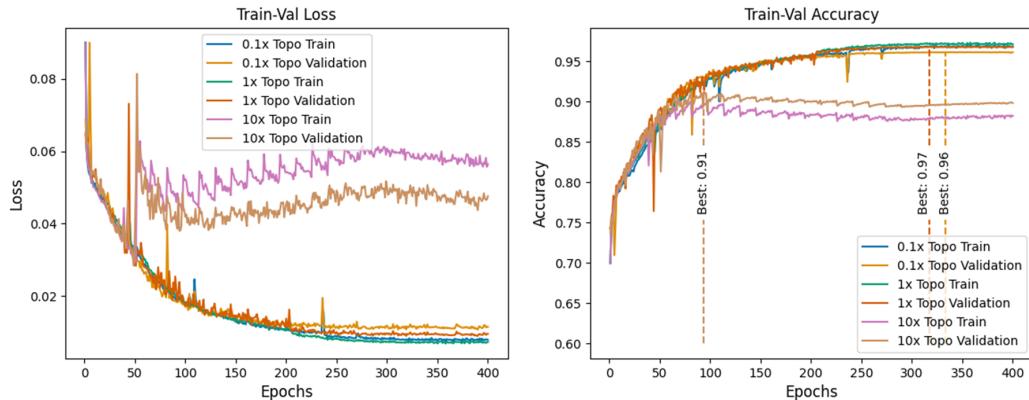


Figure 3.21: **Topological Model Training:** Composite Loss and DSC Accuracy graphs for three of the topological loss experiments with lambda values of {0.1, 1, 10} respectively. Note the best accuracy/loss values were achieved when the topological lambda factor was 1.

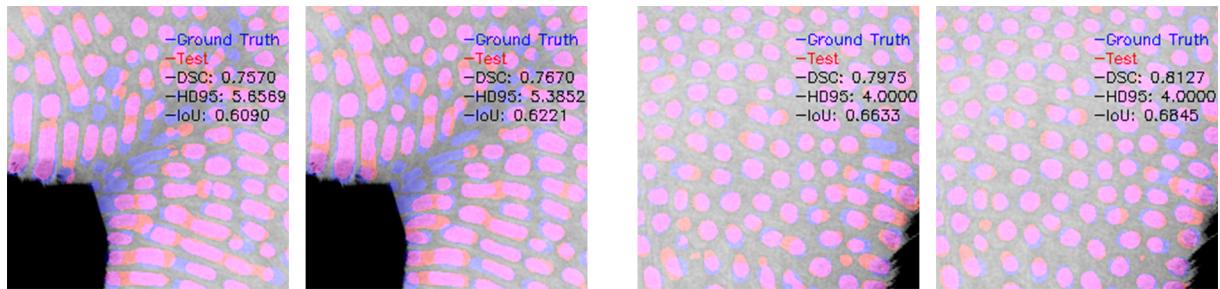


Figure 3.22: **Topological Model Inference - Unseen Axis:** Comparative composite ground-truth (blue) and prediction (red) inference outputs from the unseen Ortho testing set 3.1.1. For each pair, the left image is from the FT-VLrg model. The right image is from the same model, with the addition of warmed-up topological loss with $\lambda = 1$. Note the marginal improvement in DSC, but more importantly, the stronger topology in the region shapes, particularly notable around the folded region of the left pair. This will aid reconstruction efforts when generating the 3D model.

3.2.4 Improving Generalisation

Whilst the topological loss helped to improve generalisation on data from the same colony as the training set, inter-species generalisation was adversely affected, as issues with over-fitting became more pronounced. In order to address this, additional augmentations were introduced to the training data in the form of additive Gaussian noise, as described in Section 3.1.5.

Implementation: Three further model iterations were investigated with topological loss λ factors of 0.1, 1 and 10 respectively, with the addition of random Gaussian noise augmentations to the training set. Validation accuracies were generally unaffected, with the exception of a reduction by 1% for the model using $\lambda = 1$. This suggests that the inclusion of Gaussian augmentations to the training set has negligible effect on the model’s ability to fit the validation data. Composite training-validation graphs are presented in Fig. 3.23.

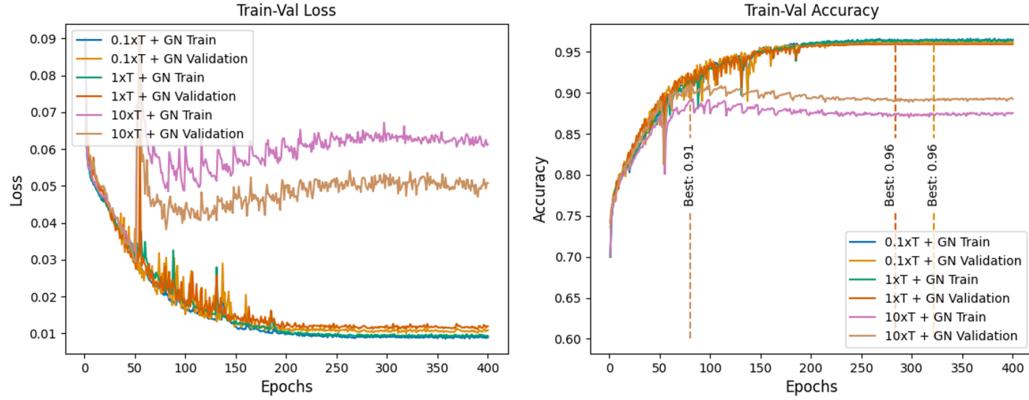


Figure 3.23: **Topological Loss + Gaussian Noise Model Training:** Composite Loss and DSC Accuracy graphs for three of the Gaussian noise (GN) experiments with topological loss (T) lambda values of {0.1, 1, 10} respectively. Note that the validation accuracy is generally equivalent to the same models without Gaussian noise (3.21).

Performance: Inference results were particularly encouraging. The Astraeiformis colony showed the best testing performance across all prior model iterations at 0.63 DSC, an improvement of +10% against the same model without the Gaussian noise augmentations. Further, a small positive increase of 0.01 DSC was seen on the unseen projection axis of the Porites 6781 slice. The known axis maintained DSC scores of 0.77. Full results are presented in Table 4.2.

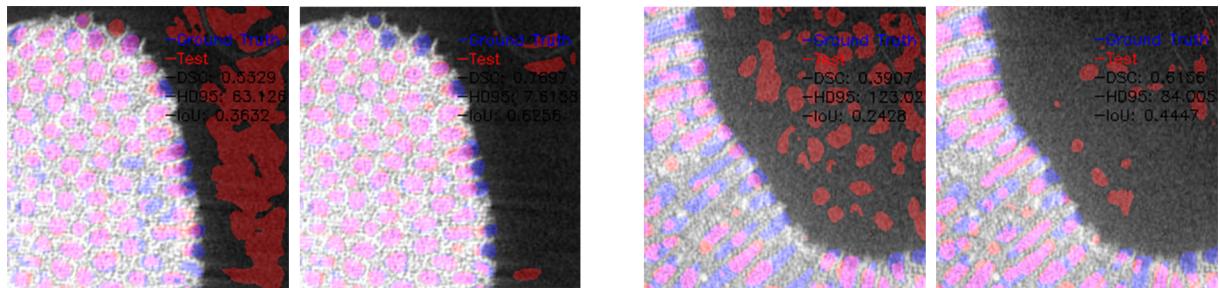


Figure 3.24: **Gaussian Noise Improvements:** Two example composite inference predictions on Porites Astraeiformis data, where blue is ground-truth and red is model prediction. For each pair the left image is predicted by a model trained without Gaussian noise augmentations, and the right with Gaussian noise augmentations. Both models are otherwise identical, having been fine-tuned using the VLrg dataset with 1x topological loss. Note, in particular, the reduction in false positives when Gaussian noise is applied.

3.2.5 Summary

This section discussed the implementation of the deep learning pipeline used for this thesis project. The model architecture was defined as a UNet-Vit hybrid with a temporal context module (TCM), which accepts volumetric inputs and generates a binary segmentation of each unique corallite region.

A naive implementation was considered, highlighting the shortcomings of the weakly annotated Naturalis 6785 dataset. A state-of-the-art (SotA) baseline model was thus established by first training a model on the weakly annotated data, and then re-training, with frozen TCM layers, using the novel annotations created by this project.

A broad variety of experiments were conducted to progress the development of a SotA corallite segmentation model. Experiments were conducted into the effects of increasing the volume of training data, and showed positive improvements to known colony predictions, at a cost to inter-species generalisation. The introduction of a custom topological loss function was established by exploring a range of λ scaling factors, and the effects of warm-up and phase-in on model performance. Finally, the addition of Gaussian noise to the training data was implemented with results suggesting significant improvement on inter-species generalisation, counter-balancing the adverse effects of previous experiments.

The thesis has now described the development of a state-of-the-art segmentation model, to predict the full range of corallite regions apparent in the provided μ CT scan data. Through the use of transfer learning techniques and a custom topological loss function, results suggest that the model is able to generate a segmentation of unseen data, to reasonable accuracy (0.77 DSC), and is able to generalise to at least one other species of the *Porites* genus, allowing for some error. This model can be used to generate a segmentation volume of an entire colony scan, providing the necessary data to automatically generate a 3D reconstruction of each individual corallite in the colony. This process is discussed in the subsequent section.

3.3 Corallite Reconstruction

Given a suitably trained model for the task of corallite region segmentation, the final phase of the project involves generating an interactive three-dimensional model of all corallites within the colony. The approach taken by this thesis is to first generate a set of segmentations from all μ CT slices of a given colony, from an arbitrary projection axis. These segmentations are stacked along the axis of projection to form a three-dimensional segmentation volume. Using the 3D modelling environment Blender⁴, ellipses are fit to each corallite region, in each slice, separated by some distance z in the environments Z coordinate system. By mapping the changes in each corallite region between slice layers, ellipses can be connected by interpolation to form a three-dimensional representation of the path and structure of each unique corallite, throughout the colony skeleton.

3.3.1 Generating Segmentation Volume

The first step of the reconstruction task involved generating predictions for each slice of the colony from a given projection axis. For brevity, the examples used throughout this discussion will be of the *Porites* Naturalis 6781 colony taken from the Growth projection axis, consisting of a total of 248 μ CT slices at resolution 4023×3487 pixels. A segmentation volume is formed by generating a segmentation prediction for each slice, and stacking these along the projection axis to form a three-dimensional data structure.

Tiling input data: In accordance with the model architecture and training environment, each slice was tiled to dimensions 224×224 , using a sliding window with step size 224, as input for segmentation prediction. The resulting output tile T is of size 224×224 where a pixel $T_{x,y}$ is assigned a value $\{0, 255\}$ based on the prediction probability $P_{x,y}$ that it belongs to a corallite region. Formally:

$$T_{x,y} = \begin{cases} 255, & P_{x,y} > 0.5 \\ 0, & \text{otherwise} \end{cases}$$

Thus, prediction tiles were mapped back to their original positions within the input, forming a full prediction slice as output.

⁴<https://www.blender.org/>

Slice Reconstruction Errors: Initial attempts highlighted an inherent problem in this approach. Given the density of corallites within each slice, each tile had partially occluded regions around the edges. This occlusion resulted in inconsistent predictions between adjacent tiles, causing visible seams in the output once reconstructed. An illustration of this is presented in Fig. 3.25.

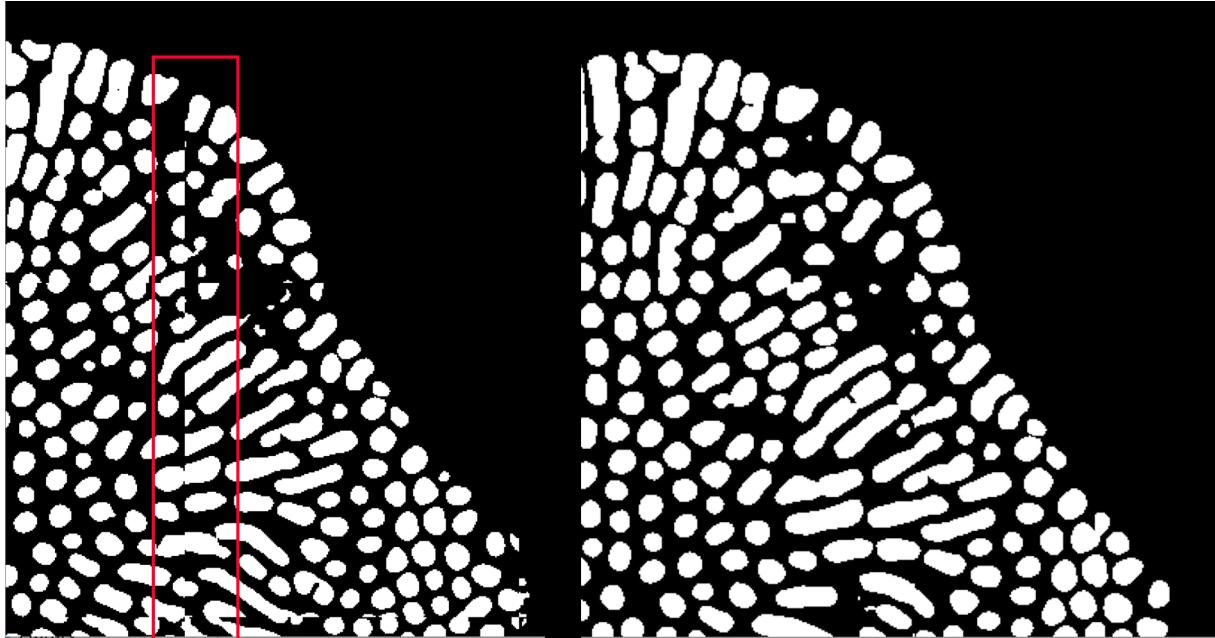


Figure 3.25: **Slice Reconstruction Issues:** An illustrative example of the appearance of visible seams when reconstructing predictions. **Left:** A subsection of a prediction reconstruction with tiling step size of 224. Note the red box identifies the regions where visible seams appear, a result of inconsistent model predictions at regions on the edge of the tile. **Right:** A subsection of the same slice with tiling step size of 112 in which no tile edges are apparent and reconstruction is ‘seamless’.

Revised Tiling: In order to address the issue of the apparent seams in the resulting slice reconstruction, the sliding window step size was reduced from 224 to 112. This generated overlapping regions between inference outputs, and by concatenating the central 112×112 window of each tile, slices could be reformed without edge seams. An illustrative example of this is presented in Fig. 3.26. This approach increases the volume of predictions required for each slice, and thus the time to generate a full colony prediction, however the trade-off is essential for improving the final accuracy of the 3D model.

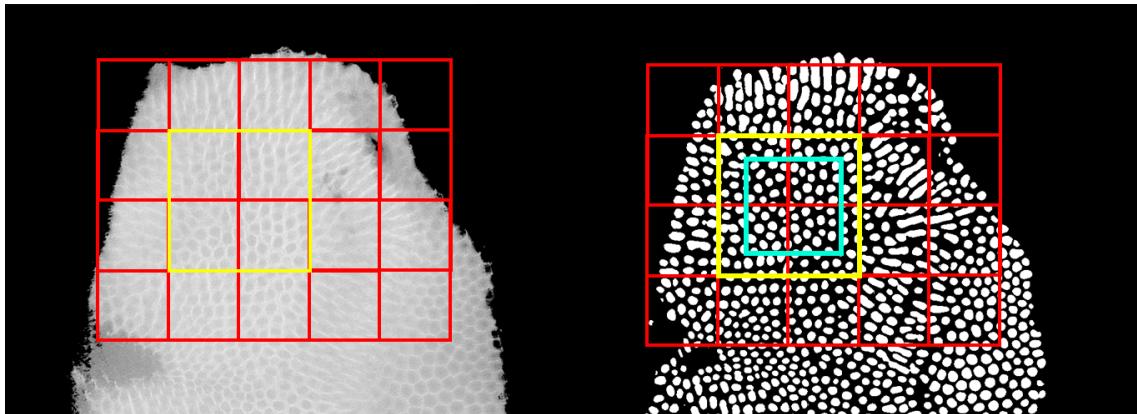


Figure 3.26: **Slice Reconstruction Example:** An illustrative example of the use of 224×224 input (yellow) with step size of 112 (red), and the corresponding reconstruction window of 112×112 (blue). This technique is used to generate seamless full-slice segmentation.

3.3.2 Corallite Tracing

In order to effectively generate a three-dimensional model, it is important to build a map of the path of each corallite through the volume by associating the region predictions in each layer with those which are adjacent along the projection axis. Although the distance between each slice is not fixed, dependent on the equipment and parameters used to capture the colony data, it is invariably small and therefore mapping regions through the projection axis can be handled through geometric analysis. It is necessary to capture some key nuances in the geometries of these mappings:

- Changes in growth direction manifest in changes in region shape, where regions will become more elongated as the growth direction becomes more perpendicular to the projection axis (Fig. 3.1).
- Polyp development, due to divergent budding or merging, results in corallites with patterns of branching and convergences throughout the colony lifespan [23].

Implementation: Given a volumetric prediction of the coral colony in the form of set of sequential slices, for each slice, each individual corallite region was identified by contiguous groups of pixels. Using the Scikit-Image package *measure*⁵ a range of region properties could be extracted, including center-points, orientations, and major and minor axis of the regions. In combination these properties can be leveraged to generate elliptical mappings to each of the regions. An example visualisation of some of these properties are presented in Fig. 3.27.

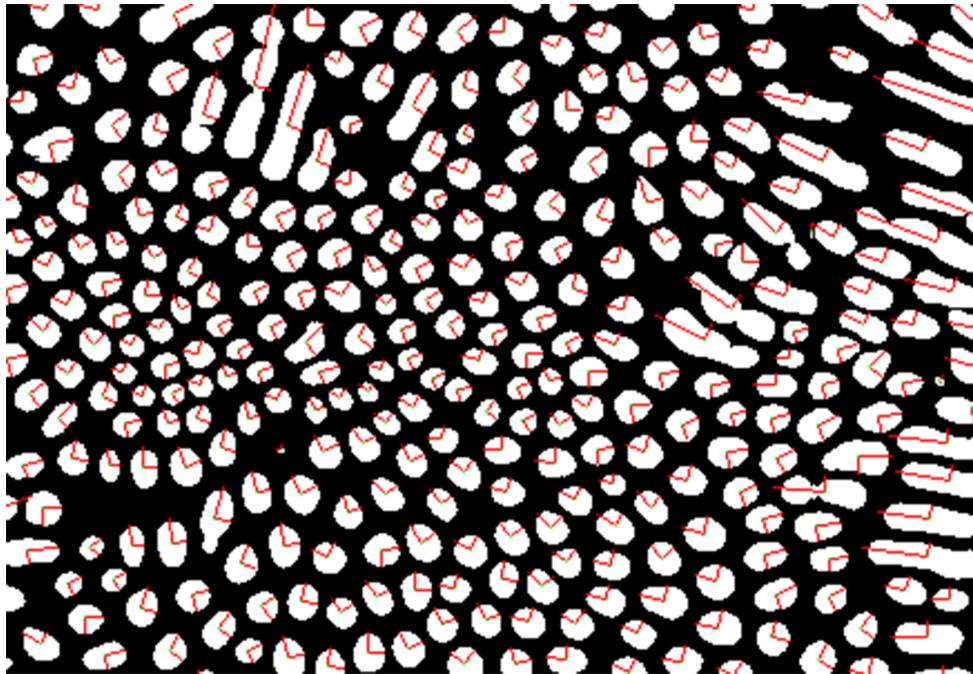


Figure 3.27: **Region Axis Properties:** A visualisation of the major and minor axis properties inherent in the skimage.measure package. These values can be used with the orientation and center pixel coordinate properties to map ellipses to each of the detected regions at a specific point in space.

Thus, beginning at one end of the volumetric prediction and working sequentially through the volume, regions at each layer could be identified as belonging to some unique corallite C_l based on a threshold function. In practise, for region R_i of slice S_n , minimum thresholds were considered for the distance of the region center from its nearest neighbour region R_j of slice S_{n-1} , and the corresponding Intersection over Union (IoU) between the two regions. If these thresholds are satisfied, R_i is added to the set of

⁵<https://scikit-image.org/docs/stable/api/skimage.measure.html>

3.3. CORALLITE RECONSTRUCTION

regions belonging to corallite C_l , of which R_j is a member:

Let:

$$\begin{aligned}
 R_i &\text{ be a region of slice } S_n, \\
 R_j &\text{ be a region of slice } S_{n-1}, \\
 R_{\alpha_{x,y}} &\text{ be the 2D center co-ordinates of region } R_\alpha, \\
 C_l &\text{ be the set of corallite regions containing } R_j \\
 d(R_i, R_j) &= \sqrt{(R_{i_x} - R_{j_x})^2 + (R_{i_y} - R_{j_y})^2} \\
 \text{IoU}(R_i, R_j) &= \frac{R_i \cap R_j}{R_i \cup R_j} \\
 \forall R_i \in S_n : &\begin{cases} R_i \in C_l, & d(R_i, R_j) < \gamma \text{ and IoU}(R_i, R_j) > \beta \\ R_i \notin C_l, & \text{otherwise} \end{cases}
 \end{aligned}$$

The values of thresholds γ and β determine the number of unique corallite regions throughout the colony model by limiting the range of 'movement' a corallite path is allowed between prediction slices in terms of center point and change in area.

Once nearest-neighbour searching has been completed, if both thresholds (γ, β) have not been met by any region in slice S_{n-1} , ie. R_i is not associated with any corallite set $C_l \in C$, a new corallite is formed: $C_{|C|+1}$.

Observations: Parameterisation of the thresholds γ and β play an important role in the structure of the corallites, in terms of the connectivity and spread attributed to a unique structure. As a result, the overall accuracy of the 3D model is dependent on these values. Trial-by-error and visual inspection suggested values of 0.3 for both thresholds yields good results, however input from a coral expert would be beneficial to refine these properties further.

3.3.3 Model Generation

At this stage of the pipeline, we have a large data set of corallite regions throughout the volume with one-to-one mappings to corallites, ie. no single region is mapped to more than one corallite structure. Further, key properties of each region have been identified and stored, providing the necessary data to build a three-dimensional model within an open-source modelling environment⁶.

The data set is typically extremely large, the number of individual corallites is dependent on the parameterisation of the thresholds discussed in section 3.3.2, but the number of total prediction regions is fixed for a given volume. The Porites Naturalis 6781 colony from the Growth projection axis consists of 714,091 regions across 248 slices. Proper optimisation of the model generation process is outside the scope of this project, although alternative techniques are considered in section 4.3.

Handling such a significant number of corallites within any modelling environment has high computational cost, with exponential growth as object complexity increases. Much discussion has preceded this section on the importance of geometric structure in the segmentation outputs. The primary motivator for this, is that it facilitates mapping of relative simple geometry (ie. ellipses) to each segmentation. By leveraging consistent simple geometry, the complexity of the model can be parameterised in terms of vertices per ellipse, which both simplifies the interpolation process when building each corallite, and gives optional detail to the end user.

Implementation: Using Blenders python-based API, an automated model generation script was implemented. The input data is formed of a list of unique corallites, each containing a set of prediction regions. A prediction region is a dictionary with the following attributes:

- Region center (x, y, z): Where z is defined by slice number * some scalar λ . For the experiments discussed in this report $\lambda = 5$.
- Region shape ($x, y, 1$): Describing the lengths of the major and minor axis of the region geometry in the x, y plane.
- Orientation (rad): Describing the orientation of the major axis around the z direction.

⁶<https://www.blender.org/>

3.3. CORALLITE RECONSTRUCTION

For each unique corallite, these properties are used to instantiate and orient an ellipse with n vertices in the x, y plane, at height z in 3D space. Each ellipse represents a region detection of the corallite, at a height proportional to the position of the slice within the segmentation volume. Once instantiated, these ellipses can be joined by interpolation of the vertices along the projection axis, represented by the z -dimension in model space, to form a cylinder-like representation of the corallite structure. An example of these stages of corallite reconstruction is presented in Fig. 3.28. Thus, each corallite is instantiated in world space as a unique object, which can be viewed in isolation, as shown in Fig. 3.29, or as a collective of any predetermined amount, as shown in Fig. 3.30.

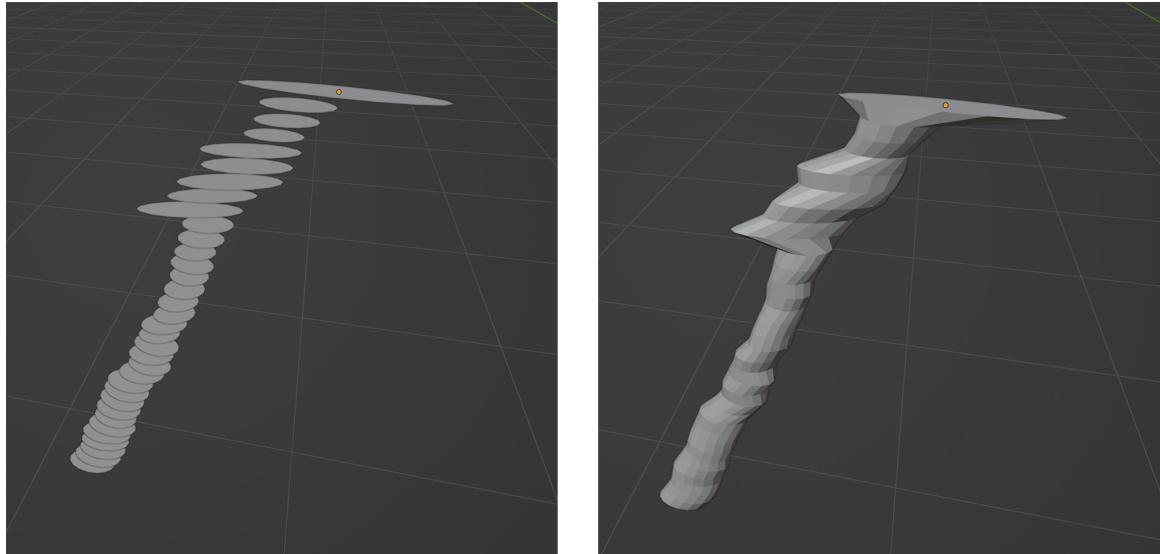


Figure 3.28: **Corallite Model Connection:** Subset of corallite #27 from the Porites 6781 Growth axis. **Left:** Instantiation of ellipses in 3D space, with 16 vertices. **Right:** Resulting corallite structure once regions have been joined.

3.3. CORALLITE RECONSTRUCTION



Figure 3.29: **Corallite Branching:** Porites 6781 Growth axis: Corallite #87 (Left) and Corallite #888 (Right) as viewed along the X-axis of the model environment. Note the curvature and branch-like behaviour captured by the implementation.

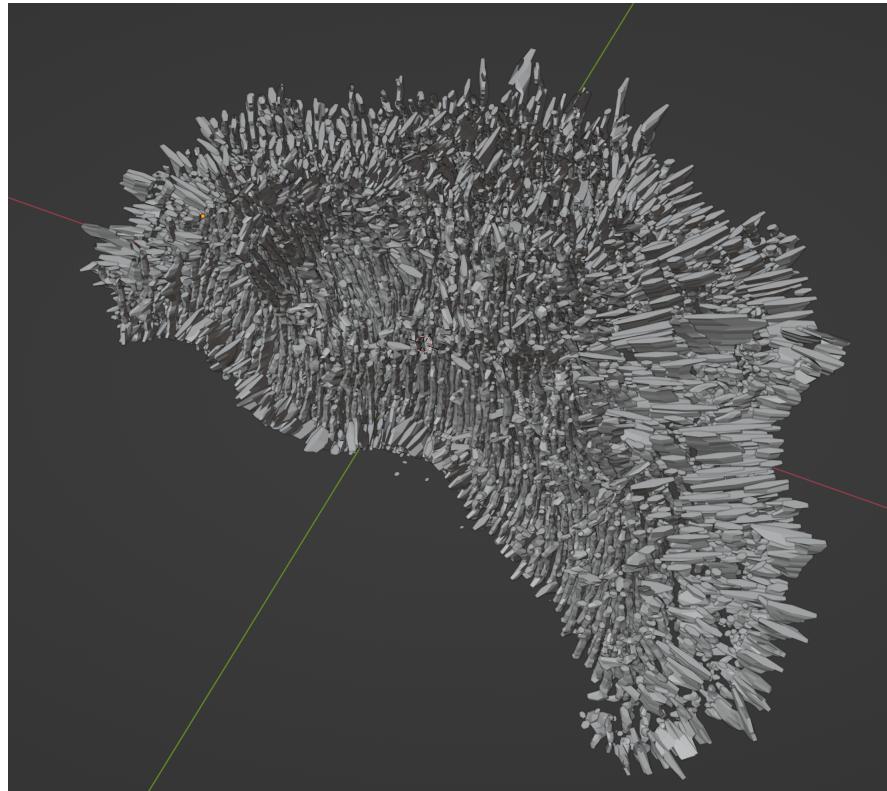


Figure 3.30: **Colony Modelling:** Porites 6781 Growth axis: Partial model of corallites #0 - #20,000 as viewed within model environment

3.3.4 Summary

This section presented the final phase of this project, concerned with automating the corallite reconstruction process. At this stage, the thesis has now described the goal of developing a complete end-to-end three-dimensional individual corallite reconstruction pipeline, as set out in section 1.1. In doing so, it has presented the first work towards such a system, and marks a significant step towards complete temporal-spatial modelling of polyp development within coral colonies of the Scleractinia order.

Given a set of volumetric corallite region predictions, unique corallites are mapped, based on the parameterised IoU and L2 norm properties of vertically adjacent regions, resulting in collections of regions along the scan projection axis which describe three-dimensional objects. These objects effectively model the branch-like behaviour apparent in corallite growth patterns.

Leveraging the Blender⁷ API, each set of regions associated to a unique corallite ID is instantiated within the model space by region-ellipse fitting, and interpolation. In doing so, an accurate model of the spatial formation of each corallite is generated. Each unique corallite can be viewed within the model space in isolation, or as a user-defined collection, with full freedom of movement in all three axis.

⁷<https://www.blender.org/>

Chapter 4

Results and Critical Evaluation

The overarching goal for this thesis is to develop an end-to-end pipe line for the three-dimensional modelling of the growth of individual corallites. The data-driven approach is supported by state-of-the-art volumetric deep-learning, with custom topological loss, and automated geometric reconstruction within an open-source graphics platform. This chapter will evaluate the implementations and results of the three key phases of the project.

First, the volumetric data generation steps will be discussed including observations on annotation deficiencies and opportunities for improvements. Subsequently, analysis of the deep-learning pipeline will be considered in terms of the three research questions set out in 1.1. The analysis will be presented from a number of angles, using the metrics defined in 2.2.4, including the performance of the custom topological loss function, and further development potential. Finally, evaluation of the model generation process will be discussed, outlining the value added by this prototype approach, and identifying areas for improvements and optimisations.

4.1 Novel Volumetric Data

Section 3.1 discussed the implementations used to generate a suitable input for volumetric deep-learning. A number of challenges were presented, and their solutions alongside any potential opportunities will be discussed below.

Annotations: The annotation set created in advance of this project was sparse in terms of the variation of corallite region geometry captured in the labels. The goals of this project set out to volumetrically predict corallite growth, ie. growth in all directions respective to the projection axis. As well as having ability to generalise between different (unseen) projection axis, and unseen species of the same genus. These aims necessitate a more comprehensive annotation set into order to train the predictive model for a broader variety of growth patterns. As such four further slices were annotated for the task, discussed in detail in section 3.1.2.

The key limitation in producing further annotations for this thesis was time, each slice took between 4-8 hours of manual labelling. A decision was made to produce a minimum viable set of annotations which could be used to train for generalisation without any overlap in the test outputs. This means that there is guaranteed separation between training and testing data, such that, testing data from the same colony is drawn from a distance away from the training data (+100 slices), or from an alternate projection axis.

Whilst these additional slices provided sufficient data to implement this thesis, there were inherent restrictions in the breadth of information available to train with. These deficiencies became particularly pronounced when testing on data from another colony, which had also been captured with different μ CT scan parameters. An example of this is illustrated in Fig. 4.1. The background (black regions) of the *Astraeiformis* scans contain a lot of noise in comparison to those of the *Porites Naturalis* 6781. The model therefore struggles to generalise well in this situation, resulting in a large number of false positive predictions. Attempts were made to address this issue by introducing further augmentation techniques, such as Gaussian noise, and are discussed in more detail in section 4.2.

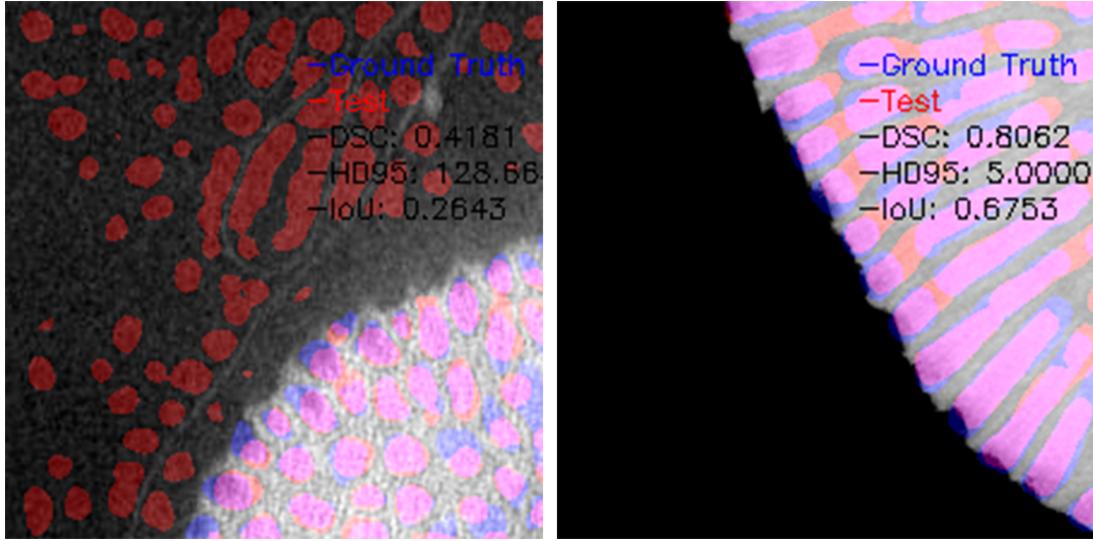


Figure 4.1: **Scan Data Inconsistencies**: Left: Composite of prediction (red) and ground-truth (blue) from the *Astraeiformis* slice. Note the high number of false positive predictions, a result of noise in the background of these slices and insufficient training data for this particular species domain. Right: *Porites Naturalis* 6781 Growth axis. Note the lack of background noise, which is consistent with the training set, resulting in no false positives in the background region.

To further improve on the potential for the model to generalise, it would be advised that a minimum set of annotations would include two slices from each projection angle of each colony, with good distance between each slice to avoid adjacency similarities. This would provide exclusive data for training, with proper separation in testing, covering multiple application domains and would likely have a dramatic improvement on the ability for the model to generalise.

Slice Tiling: Section 3.1.4 discussed the requirements and implementation of a sliding-window technique for tiling the μ CT slices prior to input to the deep-learning model. For many tasks, and in many domains, separating the input into tiles would be unsuitable, as the broader context of the original image often contains information which is pertinent to the task. There is indeed an argument for that case here. However, given the lack of annotated slices and the vast number of individual regions in each slice, there is an acceptable trade-off in order to expand volume of data available for training. Further, the contextual relationship between corallite regions is localised to a small area of the input space and there can be confidence therefore that little information is lost in tiling.

Outstanding Opportunities: Due to the broad complexity of this thesis project, there are a number of other parameters used in the data preparation stages which could be optimised through further experimentation, given suitable time. Foremost of these is the depth of snippets used as volumetric input to the deep-learning pipeline, as described in section 3.1.3. Whilst findings in [42] gives confidence that a depth of 5 was a suitable number, it would serve as interesting further exploration to find an optimal size to the training volume for this specific task. There is reasonable argument that given the close adjacency of sequential μ CT slices in the coral scans (μ m), pertinent contextual information may extend deeper than within the time-based (ms) domain of the work in [42].

4.2 Deep-Learning Pipeline

This section focuses on the evaluation of the deep-learning pipeline phase of this project, for quick reference, a description of the best model is initially summarised. The preceding discussion covers a detailed analysis of the performance of each set of experiments, both in terms of training and inference. The section will then provide responses to the three research questions set out in section 1.1 before summarising alongside a sample set of full-slice segmentations, and a full set of results (sections 4.2.5 and 4.2.6).

4.2.1 Final Model Description

This subsection acts as a quick reference point of the key parameters and configurations which formed the best model, indentified by the experiments conducted in this thesis.

- **Architecture:** As defined in section 3.2.1, with batch size 5.
- **Optimiser:** Adam, using initial learning rate of 0.0005
- **Loss:** $\frac{1}{2}(\mathcal{L}_{BCE} + \mathcal{L}_{DSC}) + 0.1 \mathcal{L}_{TOPO}$
 - \mathcal{L}_{TOPO} introduced after 50 epochs warm-up, with no phase-in.
- **Training Augmentations:** All augmentations, including Gaussian noise, as defined in section 3.1.5.
- **Datasets:** As defined under 'Fine Tuned Models' in Table 3.3.
- **Volumetric Inputs:** μCT slices tiled with a sliding window step-size of 50, and formed into volumetric structure of size $5 \times 224 \times 224$, as defined in section 3.1.3.

4.2.2 Experimental Results

This subsection will discuss the results and evaluations of the various deep-learning experiments conducted throughout the project. First, a brief discussion on the naive implementation will highlight the necessity of the additional labelled datasets provided by this thesis. Subsequent discussion will look at the state-of-the-art baseline implementation, followed by a variety of experiments exploring incremental improvements to this baseline. These improvements include expansion of training dataset, introduction of the custom topological loss, Gaussian noise augmentations, and the exploration of phase-in of the custom loss function.

Naive Implementation Prior to establishing what a state-of-the-art approach might look like, it was important to gain an understanding of the limitations of an out-of-the-box approach, and the areas in which to focus development of this thesis. The naive implementation used the existing Naturalis 6785 annotated dataset, detailed in section 3.1.1, and the VT-UNet pipeline [42], adapted for single class segmentation.

The naive model was trained using an 80% training / 20% testing split. Its performance was considered on the unseen testing data, achieving a DSC score of 0.56, presented in section 4.2.6. The resulting model predictions capture a reasonable proportion of the corallite regions described by the annotations, and even correctly predict regions which have not been annotated. Suggesting, with respect to research question DL 1, set out in section 1.1, that the underlying architecture may indeed be suitable for the task.

These predictions are, however, inherently limited by the focus of the annotations on regions of circular geometry. When considering the process of reconstructing the complete colony, it is essential to capture the full extent of corallite growth, and therefore models which leverage the novel annotations were considered in the next stage.

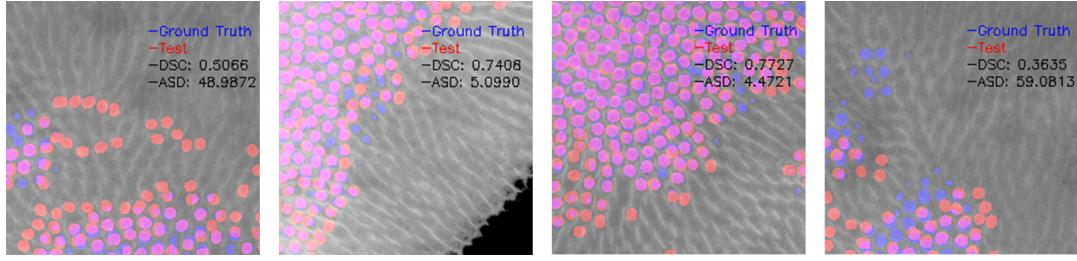


Figure 4.2: Naive Model Inference Limitations: An example set of composite prediction (red) and ground-truth (blue) inference outputs from the naive model on the Naturalis 6785 dataset (3.1.1). Note that although prediction / ground truth overlap is good, and in some cases accurate predictions have been made in lieu of ground-truth labels, a significant number of corallite regions are not captured by this model.

Fig. 4.2 illustrates the extent of the regions which are captured by this model. It is notable that even regions which are very slightly non-circular in shape are not identified as belonging to a corallite region. Non-circular elliptical geometry represents instances in which corallite growth is not aligned with the projection axis and therefore the intersection of the μ CT cross-section and the direction of corallite growth is not perpendicular. An illustration of this is presented in Fig. 4.3.

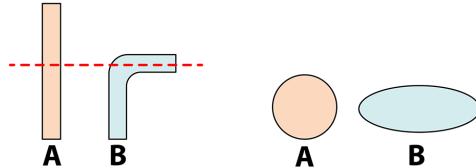


Figure 4.3: Corallite Cross-Section: An illustration of the apparent corallite geometry (right), depending on the orientation of the corallite growth with respect to the cross-section.

In the context of individual corallite reconstruction, the extent to which corallite growth can be traced is therefore severely limited using this model. This is because, throughout the lifespan of the colony, individual corallite growth is typically non-linear and thus deviates from any given scan projection axis. Identifying these limitations was instrumental in the decision to create the novel set of annotations presented by this thesis in section 3.1. For completeness, the naive model was also tested on the novel annotation set. Naturally, the model was unable to capture the annotation complexity, leading to poor performance across the board, and a maximum DSC of 0.01, as presented in table 4.2.

Fine-Tuning: A state-of-the-art baseline was therefore established for this task. The model was first pre-trained on the complete set of weakly annotated Naturalis 6785 data and then fine-tuned on the strongly annotated novel dataset, outlined in section 3.3, while keeping the TCM frozen. The motivation for this approach is that the contextual relationships between μ CT slices should be consistent regardless of the annotation strength. The weakly annotated Naturalis 6785 dataset contains a large number of annotated slices, therefore serves as a useful foundation for the model to learn these relationships. By freezing the TCM, these relationships are maintained, and the finer details of the low-volume, strongly annotated data, can be learned within other modules of the pipeline, for example, at the crucial decoding phase.

The baseline model achieved good results, with DSC scores of 0.75 0.67, 0.61 on the Naturalis 6781 Growth, Naturalis 6781 Ortho, and Astraeiformis data respectively (Table 4.2). Significantly, for the aims of this project, the novel annotations provide sufficient training data to generate a model which can predict the broader range of corallite region geometry present in the μ CT scans, used throughout this project. This is a significant step in the right direction, but introduces additional challenges in terms of ensuring that the topology of each corallite region is correct. Some baseline outputs are presented in Fig. 4.4 which illustrate the coverage of the model predictions, and highlights topological inaccuracies inside the red bounding boxes.

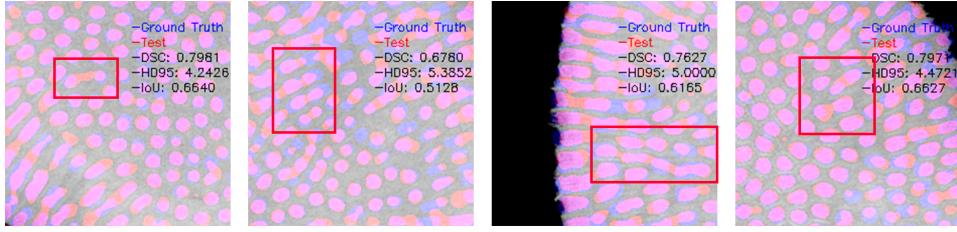


Figure 4.4: **Baseline Inference Examples:** An example set of composite prediction (red) and ground-truth (blue) inference outputs from the baseline model. Note that a much broader variety of geometry has been captured by this model, a result of the introduction of the novel annotations. Some exemplary issues with prediction topology from this model have been highlighted by a red box.

Dataset Expansion: Efforts to improve on the shortfalls of the baseline model explored expanding the volume of training data by reducing the step-size when tiling the input data prior to training (section 3.1.3). This resulted in three fine-tuned models; Fine Tuned Baseline, Fine Tuned Lrg, and Fine Tuned VLrg, using step-sizes of 224, 112 & 50 respectively. Motivating these experiments is an expectation that an increase in training data volume will encourage better corallite region segmentation, both in terms of DSC accuracy, and topological correctness.

Using a sliding window step size which is less than the size of the input to the model (ie. < 224) results in overlap in the training data, and therefore multiple instances of the same corallite region across inputs. It is expected that patterns of over-fitting will begin to emerge as the amount of overlap increases. The inference results presented in Table 4.2 certainly support this theory, where DSC increases of up to +4% are seen on the testing data from the same colony (Naturalis 6781 Growth and Ortho), but decreases of up to -5% are noted on the Astraeiformis species. This strongly suggests that the quality of inter-species generalisation has depreciated.

Visual inspection of the inference outputs show this depreciation is particularly apparent in background regions of the prediction space, resulting in numerous false positives, as illustrated in Fig. 4.5. One explanation for these issues is the disparity in μ CT scan parameters when the data is initially collected. Fig. 4.1 illustrates these differences well. As the model becomes increasingly over-fitted, it appears that the models bias begins to assign the noisy background higher probabilities that it belongs to a corallite region.

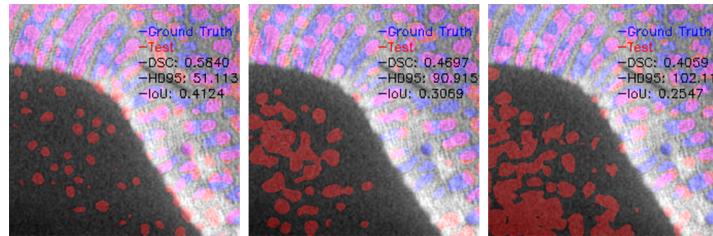


Figure 4.5: **Fine Tuning Model Inference Examples:- Astraeiformis:** An example set of composite prediction (red) and ground-truth (blue) inference outputs from each of the fine-tuned models on the Astraeiformis slice. **Left to Right:** Baseline, Fine Tuned Lrg, Fine Tuned VLrg. Note the depreciation in prediction quality is clear, particularly with respect to a significant increase in false positives in the background regions.

Visual inspection of the model predictions supports the theory that topological accuracy has been improved. This is illustrated in Fig. 4.6, where the red bounding boxes highlight areas of interest. Although improvements are made, there are still a significant number of errors. These inaccuracies are problematic when defining the geometric shapes required to generate a 3D reconstruction of the corallites later in the pipeline, and thus a custom topological loss was considered for further exploration.

Additionally, it is worth noting that the biggest DSC increase is seen in testing data from an unknown projection axis of the Naturalis 6781 colony, rather than from the same axis as the training data. One explanation for this is that the model has less room to improve on data which is closely aligned with the training set, in comparison with data from a completely unseen perspective of a known colony.

These results suggest that some measure of over-fitting has exploratory value to a potential state-of-the-art solution, particularly when improvements to generalisation of a singular colony are desired. Future improvements to these models might therefore focus on mitigating the loss in inter-species generalisation quality by other means. This will be discussed in more detail with the introduction of Gaussian noise augmentations.

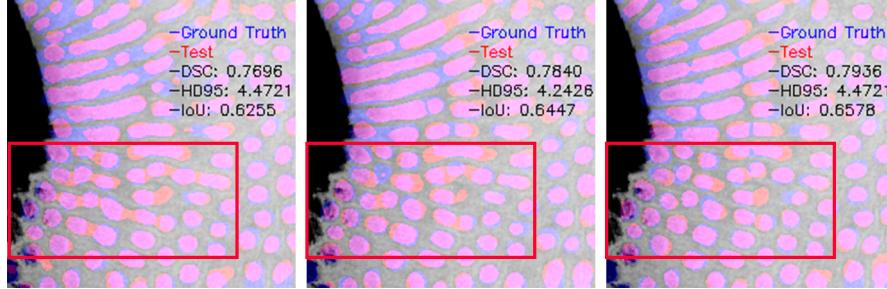


Figure 4.6: Fine Tuning Model Inference Examples:- Naturalis 6781: An example set of composite prediction (red) and ground-truth (blue) inference outputs from each of the fine-tuned models on the Naturalis 6781 Growth slice #2499. **Left to Right:** Baseline, Fine Tuned Lrg, Fine Tuned VLrg. Note the improvements in DSC metrics, and the topological accuracy illustrated by the regions inside the red boxes.

Topological Loss: The implementation of the topological loss used for this thesis has been discussed at length in section 3.2.3. One of the principle motivations for the development of this specific loss was to ensure that the topology of the model predictions was consistent, and that translation to 3D model representation was feasible. This discussion will focus on the evaluation of its introduction to the pipeline with respect to thesis goals. It is important to note that the implementation of this loss has not been optimised, particularly for GPU-based architectures, and it therefore significantly increases training time by a factor of approximately 16x.

A series of experiments was conducted to determine a suitable λ factor to apply to the topological loss. Values in the set $\{0.01, 0.1, 1, 10, 100\}$ were explored for both the baseline model, and the VLrg model. The purpose of these experiments was to understand the impacts of a custom topological loss with respect to well-separated, non-overlapping training data (Baseline), and the set of overlapping training data which has suggested performance improvements within this domain (VLrg). All experimental results are presented in Table 4.2.

Interestingly, the effects of the λ scalar are not consistent across the two sets. The baseline model shows better DSC performance with a topological $\lambda \geq 10$, whereas the VLrg model shows better DSC performance with a topological $\lambda \leq 1$. This behaviour suggests there may be a correlation between the amount of overlap in training data and the sensitivity to topological loss. The more overlap there is, the more sensitive the model becomes to topological loss, to a point where it can be detrimental to its performance. However, due to time constraints, further investigation into this theory cannot be conducted in the context of the thesis. Nonetheless, it is acknowledged as a promising area of future research to verify this hypothesis.

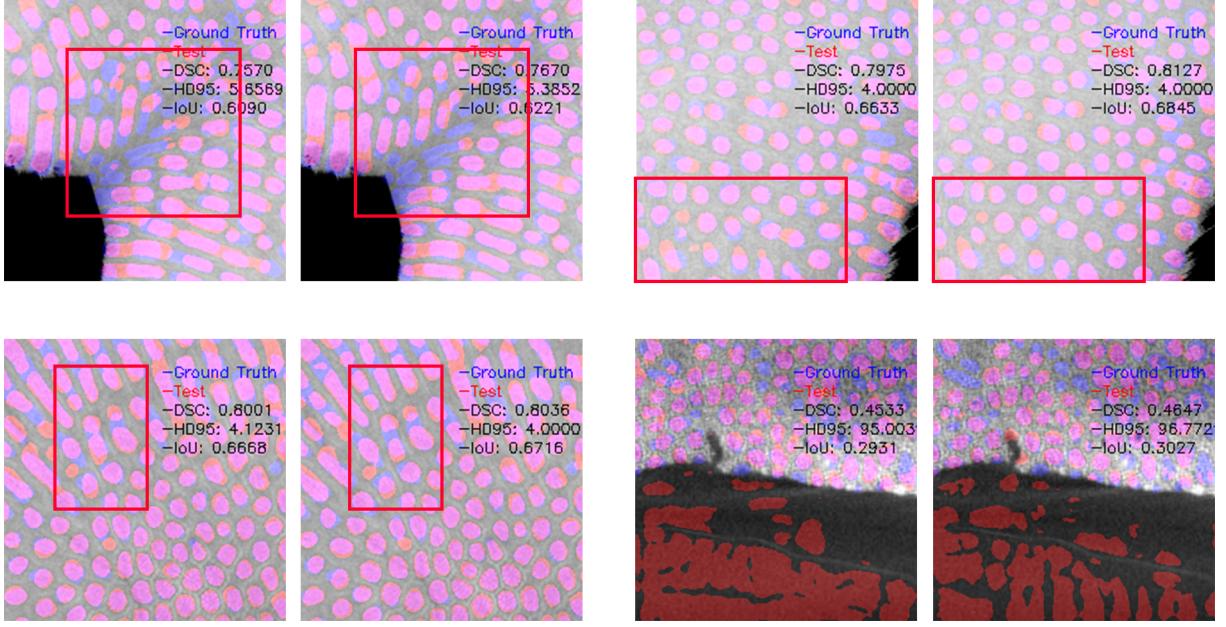


Figure 4.7: Topological Loss Inference Examples: An example set of composite prediction (red) and ground-truth (blue) inference output pairs, where for each pair the left prediction is generated by the VLRg model, and the right by the same model with topological loss ($\lambda = 1$). Note the improvement in geometric structure, particularly in the areas highlighted by the red bounding box. **Bottom right pair:** It is also important to note that topological loss has had little perceivable impact on the false positives generated in the Astraeiformis slice.

In terms of overall comparative performance between the two experimental sets, it is clear that the VLRg model with topological $\lambda \leq 1$ achieves best DSC scores on both the Naturalis 6781 Growth slice, and the Naturalis 6781 Ortho slice, the latter by at least 5%. The addition of topological loss has exacerbated the reductions in inference performance on the Astraeiformis slice, suggesting inter-species generalisation has been further diminished. For this reason, a new set of experiments was conducted to evaluate the introduction of Gaussian noise augmentations to the training data, which will be discussed subsequently. A set of example topological loss inference comparisons are presented in Fig. 4.7, where it is clear that the geometric structure of the prediction outputs are improved, as well as an observed reduction in false-positive interconnected corallite regions.

Topological Accuracy Issues: It is worth noting that the value of the topological metric developed for this thesis appears to diminish as the model performance increases. Despite clear visual improvements in the topology, in accordance with the goals of the loss, and improvements to overall DSC scores, there is little-to-no movement in the topological accuracy at inference time (Table 4.2). One plausible explanation for this behaviour is that the accuracy metric is closely aligned with the topological loss, as defined in section 2.2.4, which is non-linear by nature. Therefore, as accuracy approaches 1, the rate of change apparent in the topological metric slows.

Topological Accuracy Improvements: During the latter stages of work on this thesis, some exploration was conducted into alternative ways to measure the impacts of the topological loss, on the overall task of individual corallite reconstruction. Three additional experiments were conducted to compare the performance of the FT-VLRg model against the same model, but with the addition of topological loss. Results for these experiments are presented in full in Table 4.3.

To measure how ‘clean’ the segmentation geometry was, region compactness was considered by calculating the ratio of pixels on the perimeter of each region, to the ratio of total pixels of the region (r_p/r_a), such that, a lower ratio indicates more compact geometry. These preliminary experiments suggest an improvement of at least 3.7%, facilitating better region-ellipse fitting during construction of the three-dimensional model, as described in section 3.3.3.

Further, high-level analysis was conducted to review the impacts of this loss from the perspective of corallite region counts. Through comparison of the total prediction counts against the total ground-truth counts, the topological loss showed improvements to the count variance by at least 2%. A reduction

in the variance between total corallite predictions is not, by itself, a reliable metric to conclusively say that the topology has been improved, however it does support the hypothesis that the segmentation is more closely aligned with the ground-truth. This translates to an improvement of 3% to the total unique corallite count in the final full-colony model.

These preliminary experiments suggest some alternative ways to measure the impacts of this loss, and point to opportunities for future exploration beyond this thesis.

Generalisation improvements: In order to address the continued depreciation of performance on the Astraeiformis slice, a set of experiments was conducted to evaluate the impacts of adding Gaussian noise to the training augmentations, parameterised as described in section 3.1.5. The results of these experiments were very promising, showing positive increases of at least 10% in the DSC scores for the Astraeiformis slice, when compared with the same model without the use of Gaussian augmentations (Table 4.2). Some example inference comparisons for the Astraeiformis slice are presented in Fig. 4.8. Further, the results suggest that there is also marginal improvement on data from a known colony but unknown axis (Naturalis 6781 Ortho), and consistent performance on data from a known colony and known projection axis (Naturalis 6781 Growth).

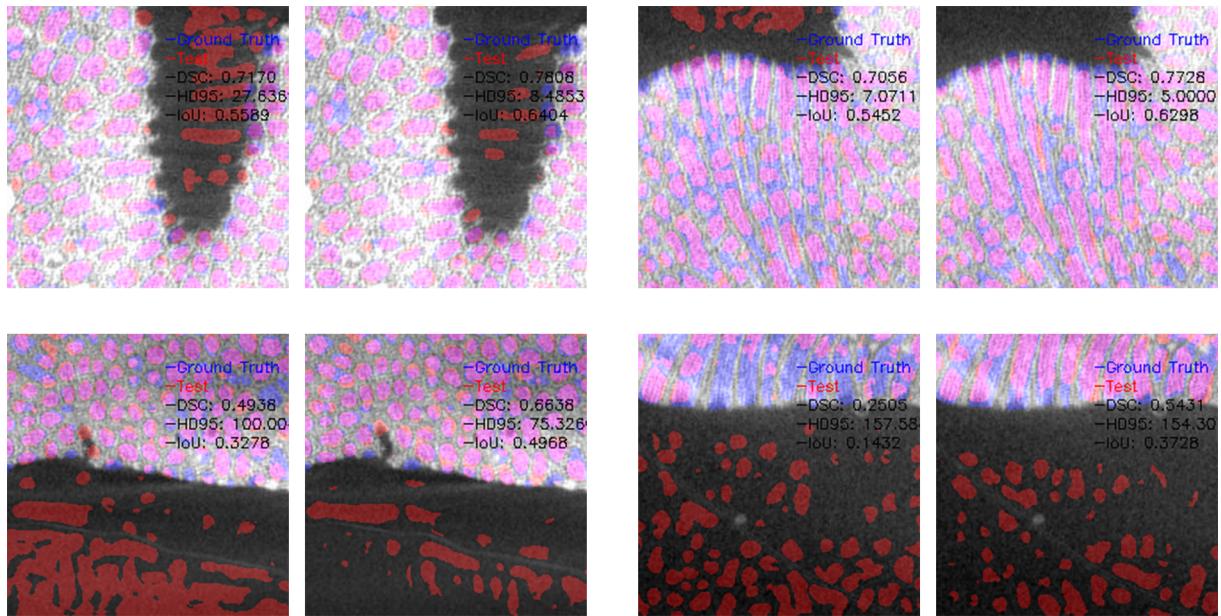


Figure 4.8: **Gaussian Inference Examples:** An example set of composite prediction (red) and ground-truth (blue) inference output pairs from the Astraeiformis dataset, where for each pair the left prediction is generated by the VLrg model with topological $\lambda = 0.1$, and the right by the same model but with the addition of Gaussian noise augmentations to training data. Note the reduction in false positives around the background regions of the tiles, an issue which had become more exacerbated by previous iterations of the fine-tuning process.

The improvements to performance on the Astraeiformis data are suggestive of improved inter-species generalisation, however, it is important to consider that the inclusion of Gaussian noise augmentations targets the differences observed between the scan quality of the training set and that of the Astraeiformis set. Specifically, it has been noted that the background of the Astraeiformis scans are noisy in comparison to the clean black background of the Naturalis 6781 scans, as illustrated in Fig. 4.1. Therefore, it is clear to say that Gaussian noise augmentations improve inter-species generalisation within the domain of the test sets used by this thesis, however it remains an open and interesting question as to whether inter-species generalisation is improved on the whole.

Phase Experiments: The final phase of experiments conducted for the deep-learning section of this thesis sought to understand the effects of phasing in the topological loss, after a pre-determined warm-up period. Three model iterations were considered using topological λ scalars of $\{0.1, 1, 10\}$. For each model a warm-up period of 50 epochs was used before linearly phasing in the topological loss according to the

following calculation, where e is the current epoch:

$$\lambda = \begin{cases} \lambda \frac{e-50}{100} & 50 \leq e \leq 150 \\ \lambda & \text{otherwise} \end{cases}$$

The results of this experiment were unexpected, despite the inclusion of Gaussian noise in the training augmentations, the inference results were reduced in comparison with the same models using a 'hard' topological loss introduction, ie. no phase-in (Table 4.2). This suggests that there might exist some relationship in the sensitivity to topological loss, and the stage of training at which it is introduced. Intuitively this is plausible, due to the non-linear nature of the topological error. Previous experiments have shown that the FT-VLrg model (with no topological loss, or Gaussian augmentations) provides good segmentation. If the topological loss is introduced, in full, nearer the point of convergence the resulting topological error will be much lower, and therefore its contribution will affect the gradient changes to a lesser extent. This hypothesis would need further investigation to validate, and is noted here as an interesting area for future study.

4.2.3 Research Questions

The goals and objectives section of this thesis set out some key research questions for the deep-learning phase of the project (1.1). This subsection will discuss the findings of the experiments with respect to those key questions.

DL 1: To what extent does a volumetric segmentation model, designed for μ CT data in the medical imaging domain ([42]), translate to the domain of μ CT corallite segmentation?

It is important to first note that there is, by nature of engineering, at least some minor adaptation required to translate a system from one domain to another. In the context of this thesis, this required modification of the existing pipeline from a multi-class segmentation system to a single-class segmentation system. Further, a evaluation of the model performance using strongly, and weakly annotated data has shown it is imperative that the training data is representative of the requirements of the task.

By introducing a set of novel annotations with comprehensive corallite region coverage, and fine-tuning a model trained on a large volume of weakly-annotated data, with this strongly annotated dataset, these conditions can be met.

Results are evident, even at the baseline, that domain transfer is straightforward and produces good results, of at least 0.75 mean DSC, compared to the target annotations. This suggests that the underlying UNet-ViT architecture first proposed by [5] has general purpose application for segmentation tasks, at least within the domain of micro-computed tomography imaging. Unsurprisingly, given the close relationship between time and space, these results also suggest that the TCM introduced by [42] is suitable for encoding relationships in either domain. Exploration of the importance of the TCM in the context of corallite reconstruction would, however, be an interesting ablation study for future work. This would help understand the extent of the volumetric relationships in the μ CT data and support establishing an optimal depth encoding for the task of corallite reconstruction.

DL 2.1: To what extent does the introduction of a custom loss function improve on a baseline state-of-the-art segmentation solution?

Through the process of implementing and evaluating this loss, several relationships have become apparent, pertinent to answering this question. Discussed in detail in section 4.2.2, the results suggest correlations between the effect of the topological loss and the amount of overlap in the training data, and how close the model is to convergence when the loss is introduced. The investigation of these relationships would be necessary to answer this question fully. However, in the context of the model iterations explored by this thesis, in purely metric terms, the introduction of the topological loss provided a minimum 1 – 2% DSC increase across the three test sets, as can be seen by comparing the best model (bold) and the model above, with topological $\lambda = 0$ in the results presented in Table 4.2.

An important note for discussion is the added computational complexity introduced by this loss. In its current form, which has not undergone any form of optimisation, particularly for GPU deployment, training times are increased by a factor of approximately 16. This time is significantly detrimental to the ability to iterate and evaluate the effects of this loss in great detail and raises questions to its performance-cost trade-off value in its current non-optimised form.

DL 2.2: To what extent does the introduction of a custom loss function support the generation of a 3D representation of unique corallites?

One of the key properties of what a good segmentation looks like, in the context of the automated 3D corallite modelling system outlined in section 3.3, is well-formed geometry that facilitates fitting simple ellipses to the segmentations. For this reason, the topological loss specifically targets the false-positive regions between two, or more, separate corallite labels. As discussed in section 4.2.2, the topological accuracy metric does not provide the necessary detail to describe the fine-grained changes at this level, and therefore providing a quantitative answer to this question is somewhat more challenging. It does show improvements on the Astraeiformis slice, where the scope for change is greater, and we can say that the topological accuracy in the other test domains is at least as good, according to this metric. Additional means to measure success have been proposed and preliminary results presented in Table 4.3 support the hypothesis that the topological loss improves corallite count errors by at least 2% and compactness of geometry by at least 3.7%.

From a qualitative perspective, the effect is much clearer. Examples presented in Fig. 4.7 are representative of the kinds of adjustments in inference quality observed by this loss. Noting that although the measured differences are small, the objective of improving region separation with cleanly defined geometry, as motivated in section 3.3.3, is well achieved.

Therefore, we can say that there is qualitative evidence that the custom loss function introduced by this thesis improves the region separation and geometric definition in the segmentation outputs. These regions of improvement are crucial properties of a suitable input for the automated modelling system. There is little evidence to suggest that these improvements are circumstantial, however, future work might look to improve on the topological accuracy metric to further validate the effects of this loss in a more concrete, quantitative manner.

4.2.4 Summary

This section has evaluated the deep-learning pipeline phase of the project. A thorough evaluation of each of the key model iterations has been discussed, showing incremental improvements throughout the experimentation process. The evaluation process has observed the importance of annotations which provide full coverage of the domain, and the value of using a limited quantity of strongly annotated data to improve on a model trained with a large quantity of weakly annotated data.

The effect of increasing the volume of training data, using overlap, has been shown to have a positive effect on predictions of unseen data from a known coral colony, but a negative effect on the ability to generalise between different species. These are properties of over-fitting, however, they can have value to the final model performance if mitigated by other means, as has been suggested by the experiments into Gaussian noise augmentations.

Finally, the introduction of topological loss has been shown to have a positive effect on the quality of corallite region separation and geometric consistency across all testing domains. Although it is noted that the topological accuracy metric is unable to well-describe the fine-grained topological changes observed by qualitative evaluation, alternative methods by which to measure the impacts of the topological loss have been suggested, with preliminary results presented in Table 4.3.

Throughout the implementation and evaluation of the deep-learning pipeline a number of further questions and possible routes of future investigation have been raised. A summary collection of these are provided in chapter 5. A random selection of full-slice segmentations are presented next, and a full set of experimental results presented at the end of this section. The final section of the evaluation chapter will follow, concerned with the automated 3D corallite reconstruction phase of this thesis.

4.2.5 Full Segmentation Examples

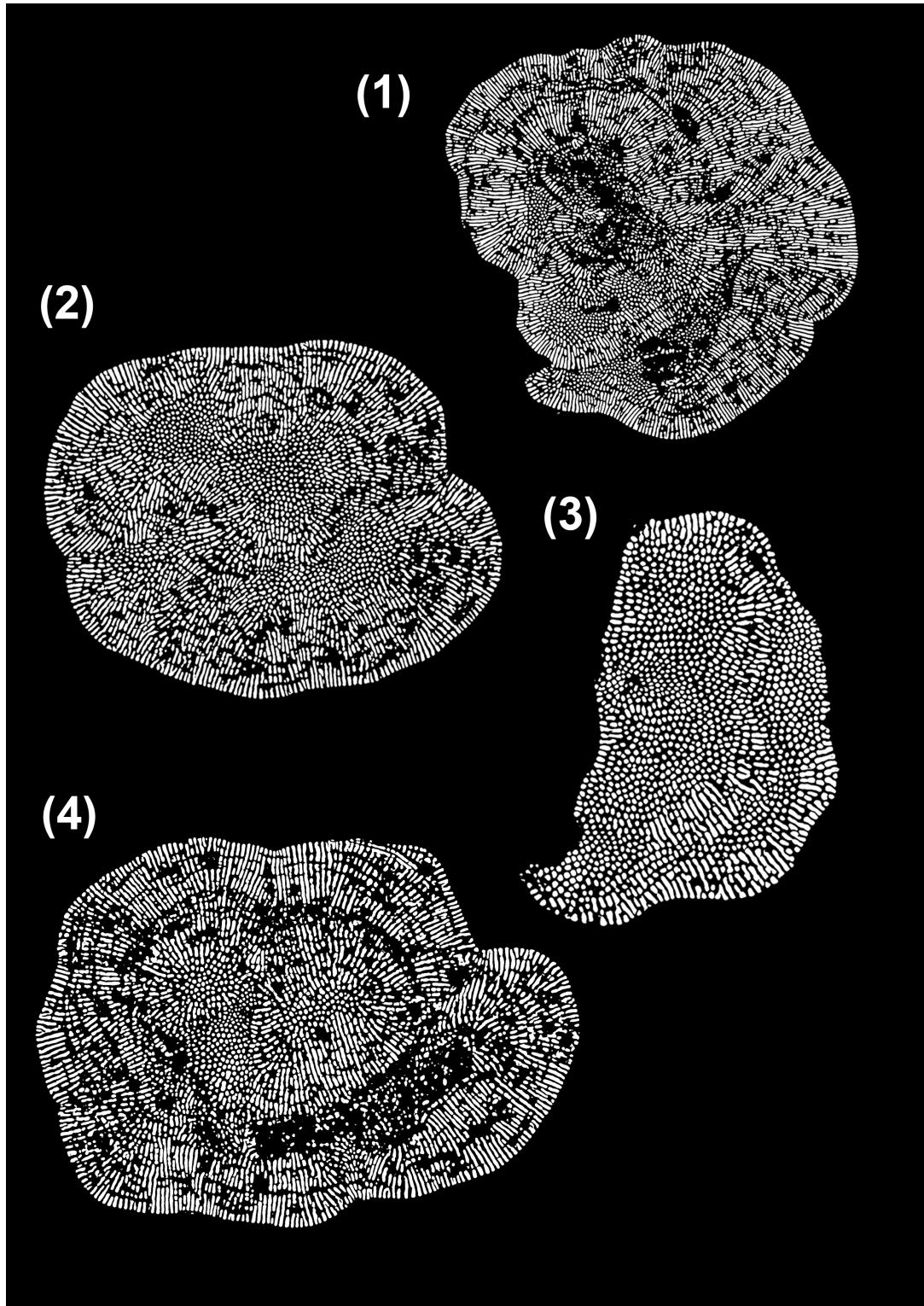


Figure 4.9: **Complete Slice Segmentation's:** A randomly selected set of segmentation slices from the Naturalis 6781 colony scans, generated by the optimal model identified by this thesis, outlined in section 4.2.1. (1): Growth slice #1919, (2): Ortho slice #2368, (3): Growth slice #1194, (4): Ortho slice #2078. Note that, in general, the model achieves excellent separation between corallite regions, and is able to describe a range of region geometry representative of the omni-directional corallite growth patterns. This is crucial for the success of the corallite reconstruction phase of this project.

4.2.6 Results

Model	Test Set	
	Naturalis 6785 (Test)	
	DSC	Topo
Naive	0.56	0.95

Table 4.1: **Naive Model:** Performance of the Naive model on the Naturalis 6785 testing set. The Naturalis 6785 data is unsuitable for performance analysis of fine-tuned models due to the sparsity of annotations, however the results are indicative of performance when applied to the same data domain as the training set. **DSC:** Mean Dice similarity, **Topo:** Topological accuracy as defined in 2.2.4.

Model	Test Set					
	Naturalis 6781 Growth #2499		Naturalis 6781 Ortho #1003		Astraeiformis	
	DSC	Topo	DSC	Topo	DSC	Topo
Naive	0.008	0.66	0.01	0.75	0.01	0.68
Fine Tuned Baseline	0.75	0.93	0.67	0.92	0.61	0.88
Fine Tuned Lrg	0.76	0.94	0.69	0.93	0.57	0.80
Fine Tuned VLrg	0.76	0.94	0.71	0.93	0.56	0.79
FT Baseline 0.01 Topo	0.76	0.93	0.65	0.92	0.59	0.85
FT Baseline 0.1 Topo	0.76	0.93	0.64	0.92	0.59	0.85
FT Baseline 1 Topo	0.76	0.93	0.62	0.90	0.60	0.85
FT Baseline 10 Topo	0.76	0.93	0.65	0.92	0.59	0.85
FT Baseline 100 Topo	0.77	0.93	0.62	0.90	0.60	0.85
FT VLrg 0.01 Topo	0.77	0.94	0.70	0.93	0.51	0.80
FT VLrg 0.1 Topo	0.77	0.94	0.70	0.93	0.52	0.80
FT VLrg 1 Topo	0.77	0.94	0.70	0.93	0.51	0.79
FT VLrg 10 Topo	0.76	0.94	0.69	0.93	0.48	0.80
FT VLrg 100 Topo	0.76	0.94	0.69	0.93	0.53	0.81
FT VLrg 0 Topo + GN	0.76	0.94	0.70	0.92	0.61	0.87
FT VLrg 0.1 Topo + GN	0.77	0.94	0.71	0.93	0.63	0.87
FT VLrg 1 Topo + GN	0.77	0.94	0.70	0.93	0.60	0.86
FT VLrg 10 Topo + GN	0.77	0.94	0.70	0.93	0.63	0.89
FT VLrg 0.1 Topo (P) + GN	0.76	0.94	0.70	0.93	0.48	0.78
FT VLrg 1 Topo (P) + GN	0.77	0.94	0.70	0.93	0.53	0.81
FT VLrg 10 Topo (P) + GN	0.77	0.94	0.70	0.93	0.54	0.84

Table 4.2: **Inference Results:** Full testing performance for all model iterations. The model in bold is selected as it offers the best, or joint best, performance across all testing domains, with the exception of the topological accuracy on the Astraeiformis data. Naturalis Growth #2499 is an unseen slice from the same colony and same projection axis as the training data. Naturalis Ortho #1003 is an unseen slice from the same colony but different projection axis as the training data. Astraeiformis is an unseen slice from a different colony, and species, of the Porites genus.

x Topo: The lambda factor applied to the topological loss function. **(P):** Phase-in of topological loss over 100 epochs. **+ GN:** Addition of Gaussian noise to training data. **DSC:** Mean Dice similarity, **Topo:** Topological accuracy as defined in 2.2.4.

Metric	Model		Topo Improvement
	FT-VLrg	FT-VLrg + 0.1 Topo	
Mean Compactness (perimeter/area)	0.28	0.27	3.7%
Corallite Count Error (total, full slice)	0.1	0.08	2%
Unique Corallite Count (final model (less is better))	215298	208725	3%

Table 4.3: **Alternative Topological Accuracy Metrics:** Alternative metrics used to support the hypothesis that topological loss improves topological accuracy. Testing was conducted on the Naturalis 6781 Ortho dataset 3.1. **Compactness:** Measures how ‘clean’ the geometric structure of each region is. A lower perimeter-to-area ratio suggests tighter geometry, which is preferable for the region-fitting tasks when generating the three-dimensional reconstruction. **Corallite Count Error:** The error in total number of corallites predicted in a complete slice. **Unique Corallite Count:** The total number of unique corallites counted in the final full-colony model.

4.3 Corallite Reconstruction

This section focuses on the evaluation of the automated corallite reconstruction phase of the project. The principle goal of this phase, and the overall project, is to produce a 3D representation of the historic growth patterns of each unique corallite throughout the colony. Therefore, given a set of segmentation predictions generated by the deep-learning pipeline, unique corallites are first identified, and then mapped through the segmentation volume. This mapping is translated into geometric object data using ellipses parameterised by vertex count, and interpolation along the projection axis.

The final model is a set of many thousands of unique corallite objects which can be selectively viewed in isolation or as groups within the model environment, where full camera orientation and position control is available.

It should be noted that complete development of this phase is a software engineering project in its own right, and as such, this thesis offers a prototype approach. In doing so, and in combination with the deep-learning pipeline established prior, it takes the automation of 3D individual corallite reconstruction a significant step towards a complete spatial-temporal solution. Section 3.3 detailed the implementation of this phase in terms of three steps, used again here to guide the evaluation process.

Generating Segmentation Volume: The process of constructing the segmentation volume involves running inference on all slices of the colony from the chosen projection axis, this generally involves hundreds of high resolution images. The implementation described in section 3.3.1, involves tiling each slice to dimensions 224×224 , and then reconstructing the full slice from the segmentations produced by the deep learning model. The key challenge in this process was the evidence of seams in the reconstructed slice. It was observed that around the edges of the input tiles, corallite regions were incomplete and thus partially occluded. At inference time this generates inconsistent predictions between adjacent tiles, and when combined, the whole corallite region is incorrect or distorted, resulting in visible seams. The edges of each segmentation are therefore shown to be unreliable.

In order to ensure a seamless reconstruction of the whole slice, overlap in the inference tiles was created such that only the central 112×112 region would be used for reconstruction. The illustration and implementation of this can be reviewed in section 3.25 and some randomly selected example segmentations are presented in Fig 4.9. This process increases computational complexity of the inference process, as the volume of data is increased. However, this trade-off is necessary, the prototype system developed for automating the 3D reconstruction relies on fitting ellipses to well-formed geometry, which is significantly inhibited by the presence of seams.

Once correctly reconstructed, each segmentations slice can be stacked along the axis of projection, such that the μ CT scan sequence is preserved, thus forming a segmentation volume of the corallite regions. It is important to note there is some unknown distance, at μm scale, between the raw μ CT slices. This distance, in combination with the likelihood of some inaccuracies in the segmentation outputs, means that mapping the corresponding segmentations through the volume is a non-trivial process.

Corallite Tracing: Given a volume of segmentations, as defined by the previous step, each unique corallite requires identification throughout the extent of the volume. Every corallite region, in each slice, must be therefore mapped to a unique corallite structure. In addition to this, suitable properties of each corallite region were extracted to facilitate effective modelling of the corallite within the 3D environment. This is discussed in detail in section 3.3.2.

Region center coordinates and major and minor axis are key values, with respect to ellipse mapping within the model space. Additionally, co-ordinates of a bounding box around each region were stored to enable an approximate IoU calculation between sequential layers of the volume.

This data enables a process which maps regions in slice n to regions detected in slice $n - k$, through parameterised comparison of the L2 norm of the region centers, and the IoU of the approximate bounding boxes. These properties are by no means comprehensive, however computational demands to map over 700,000 regions are already high, and for the purposes of this thesis, ensuring this process was simplified facilitated broader exploration of the solution. There are several additional properties which are noted for possible future exploration, such as the orientation of ellipses, to restrict the range of rotation the modelled region is permitted, and nearest neighbour overlap, to mitigate issues by which a regions from neighbouring corallites are assigned to the wrong structure.

The mapping algorithm described in section 3.3.2 guarantees that each region in a slice is mapped to one existing corallite from the previous slice, or is instantiated as a new, unique corallite. Therefore, the parameterisation of the thresholds for the L2 norm and IoU comparisons are integral to the connectivity properties of the resulting corallite mapping. If the threshold is too low, the resulting corallite structure contains many branches, in what appears to be multiple, true corallite pathways, mapped as one. If the threshold is too high, region connectivity is poor, resulting in what should be a single corallite structure, presenting as several instances of short, detached corallite structures. The parameterisation of these thresholds is an act of balancing these conflicting behaviours, and would ideally require discussion and analysis with an expert in corallite growth patterns. Fig 4.10 provides a visual illustration of these behaviours.

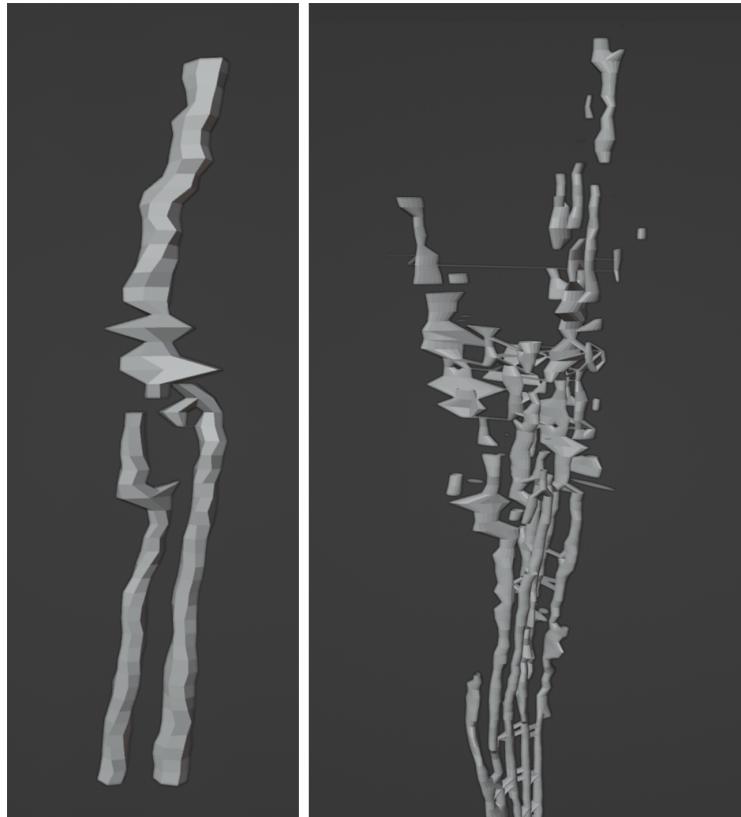


Figure 4.10: **Model Connectivity Behaviours:** Illustrative examples of adversity in corallite connectivity behaviours in relation to thresholds of IoU and L2 norm when tracing corallite paths. **Left:** Threshold values are too high, resulting in separate corallite structures, where one should be formed as two corallites merge. **Right:** Threshold values are too low, resulting in one corallite structure, where there should be several.

The implementation used by this thesis compares the properties of regions in slice n , with the properties of regions in slice $n - 1$, however, there are outstanding issues in connectivity. For example, the segmentation output in slice $n - 1$ may have segmentation error, resulting in a missed region. This translates to the formation of a new unique corallite at the height of slice n . The resulting model now suggests two separate corallites, as shown in Fig 4.11. An improvement to this implementation might consider comparisons with regions in slice $n - 2$, and further, in cases in which a suitable existing corallite structure has not been identified. However, it is important to note that without optimisation, this will increase computational time.



Figure 4.11: **Missed Regions:** An illustrative example of regions missed by the segmentation model resulting in separate corallite structures. When a region at height n is unable to find a region at height $n - 1$, which meets the parameters of the search described in section 3.3.2, a new corallite structure is formed. This example shows five 'unique' corallites, where there should be one. Extending the depth of the search for region-corallite mappings would help to further mitigate these issues.

Model Generation: This stage of the corallite reconstruction phase involved translating the corallite mappings described above into geometric object data. This was achieved using the python API for Blender¹. Having python-based access to Blender, alongside the comprehensive documentation², significantly helped to expedite implementation of this step. Naturally, there are nuances to the way in which certain operations are implemented in the API, particularly with regards to the interpolation between corallite regions that split and merge, and the implementation provided by this thesis is by no means optimised for computational efficiency.

The key goals for this step were to ensure that each corallite region was instantiated as an ellipse in world space with the appropriate position, scale, and orientation. Further, it was necessary to ensure that interpolation between ellipses, along the z-axis, occurred only between regions associated with the same unique corallite. The approach was therefore to build each unique corallite one-by-one, by instantiating all regions as ellipses, and interpolating along the axis of projection, represented by the z-axis. The resulting model space is thus formed of a collection of n objects, representing the n unique corallites modeled by the project pipeline.

This particular approach ensures that each corallite can be identified by a unique object name, and can be interacted with as an individual, or as part of a specified collection within the bounds of the 3D environment. This includes full control of camera position, orientation, and zoom, providing a meaningful way to observe the properties of each corallite model.

There are some notable properties to this implementation, and the Blender API, which are worth further discussion. Of primary note is the execution time to generate a model for colonies consisting of large numbers of corallites. The process to build each corallite model is fairly protracted as it requires switching the program states of Blender, in order to access certain parts of the API. The result of this is that it takes approximately one hour to model 1000 corallites on a 2.3Ghz QuadCore i7 Macbook Pro. The total number of unique corallites is dependent on the parameterisation of the corallite tracing process, but for all colonies tested by this thesis, the number is invariably in the tens of thousands. This raises questions as to whether there are alternative representations inside the 3D environment, which might be computationally more efficient to build.

One avenue that is noted for potential exploration is the mathematical representation; nonuniform rational B-splines, or NURBS, which solve for curves between two points. There are limitations to the level of complexity NURBS can effectively model, particularly if the system is automated and manual adjustments are not an option, but there are vastly fewer data points and therefore model generation is likely to be exponentially faster. An interesting exploration for future work might seek to understand the differences in model detail, representation, and generation efficiency between the vertex-based approach this thesis takes, and one which uses NURBS with region center points, for example.

Corallite Reconstruction Summary: Section 1.1 set out some of the key aims for the prototype automated model generation system explored by this thesis. Central to these aims was the ability to model corallite development in terms of unique structures, which could be observed and analysed in all three-dimensions. The system presented gives proof-of-concept of a parameterise-able approach which facilitates 3D corallite reconstruction. Users can observe individual corallites, selective groups, or the whole colony will full freedom of movement. With further work towards optimisation and in collaboration with experts in corallite development, the system presented shows potential to be an effective path towards a full spatial-temporal reconstruction.

¹<https://www.blender.org/>

²<https://docs.blender.org/api/current/index.html>

4.4 Results and Evaluation Summary

This chapter has evaluated the project pipeline in terms of the development goals, experimentation results and research questions set out in section 1.1.

First, the volumetric data generation phase was discussed, where the strongly annotated novel dataset presented by this thesis was shown to significantly improve the foundations on which the project pipeline was built, by expanding prediction coverage to all corallite regions. Guidelines on minimum annotation sets were presented, suggested for future improvements on model generalisation, or to be considered if exploring the transferable properties of the model to other genera of coral. Opportunities were noted in further exploring the contextual relationship of corallite regions along the projection axis by mapping the correlations between snippet depth and inference performance.

The following section extensively covered evaluations of the deep-learning phase of this project. An evaluation of the naive implementation showed the inadequacies of training with only the weakly annotated Naturalis 6785 dataset, thus reinforcing the value of the novel datasets. A state-of-the-art baseline was presented, showing immediate performance strength and suggesting that, in answer to question DL 1, the underlying architecture develop by [42] transfers well into other domains. Experimental iterations were shown to progress inference performance on the baseline, on data from completely unseen slices on known colonies, through both the expansion of training data by overlap, and the introduction of a custom topological loss. These improvements had adverse effect on inter-species generalisation, however, the addition of Gaussian noise augmentations at training time was shown to significantly rectify this within the domain of the testing datasets.

Finally, the prototype automated reconstruction phase was considered. The implementation explored by this thesis was shown to generate individual 3D corallites within an environment which facilitated full freedom of exploratory movement, and therefore structural analysis in the spatial domain. Noting that the implementation is sub-optimal in terms of performance, and that temporal development is, at present, an approximation in relation to height, this system provides evidence that this approach takes a significant step towards a complete automated spatial-temporal modelling of all individual corallites of a Porites colony.

Chapter 5

Conclusion

The final chapter of this thesis will be divided into two sections. The first section will provide an overview of the project status, summarizing the contributions and responses to the goals and questions outlined in section 1.1. The second section will focus on future opportunities, exploring the various avenues for further investigation that were identified during the project's execution, evaluation, and analysis.

5.1 Project Status

This section will comprise a summary of the main contributions followed by a description of the current state of the project with respect to the goals set out in section 1.1.

Contributions: Throughout the exploration and delivery of this thesis, the following novel contributions have been made. In doing so, the work presents the first description known to the author and his supervisor of a complete pipeline for automated three-dimensional individual corallite reconstruction. This work, and its contributions, mark a significant step towards complete spatial-temporal colony reconstruction:

- The creation of a novel annotation set of μ CT slices from colonies of the Porites genus. Annotations for each slice provide comprehensive coverage of all corallite regions. In total, 8412 new corallite region labels are provided across 4 slices from the Porites Naturalis 6781 dataset, and the Porites Astraeiformis dataset.
- The implementation and evaluation of a state-of-the-art TCM-UNet-ViT-based deep learning pipeline, built for volumetric corallite region segmentation with micro-computed tomography data.
- The introduction of a topological loss function designed to enhance segmentation topology in the context of high density ellipse detection. Noting that there may be some transferability of such a loss to other domains, for example, cell segmentation in the field of medical imaging.
- A prototype automated 3D corallite reconstruction system, providing evidence for a viable interactive spatial-temporal model of all individual corallites of a Porites colony.

Goals and Objectives: This subsection will provide a summary of the findings and observations in relation to the key goals, objectives and research questions, which this thesis set out in section 1.1.

The first phase of the project considered, and presented, a variety of techniques used to form a suitable volumetric input for the deep-learning architecture used in the second phase of the project. A novel set of strongly annotated corallite region data was presented, and through evaluation it was shown that these were crucial to the success of training for this task, and ultimately of the overall project. The necessity of reducing the input dimensionality through tiling was discussed, and the subsequent construction of three-dimensional snippets from raw μ CT data was described. Augmentations employed in the training pipeline were detailed and later shown, particularly in the case of additive Gaussian noise, to play a significant role in improving inter-species generalisation performance.

The second phase of the project was concerned with the implementation and evaluation of a deep-learning pipeline which could produce a comprehensive and viable segmentation of the entire μ CT volume. It was important that the segmentation was produced not only to a reasonable degree of accuracy, but provided suitable region separation to facilitate three-dimensional individual corallite reconstruction. This phase was explored in the context of three key research questions which have been summarised below, however detailed discussion is covered in section 4.2.3;

- **DL 1: To what extent does a volumetric segmentation model, designed for μ CT data in the medical imaging domain ([42]), translate to the domain of μ CT corallite segmentation?**

The work in this thesis has shown, within the bounds of the datasets it explored, that the architecture described by [42] transfers from the domain of medical imaging to domain of μ CT corallite segmentation very well. This is conditional, as with all DL tasks, on ensuring that the labelled data is fully representative of the task. Leveraging the novel annotations, a state-of-the-art baseline was defined, needing little architectural adaptation, and delivering impressive results. Despite improvements made in the latter stages of this phase, the baseline implementation evidences the transfer-ability of the model architecture.

- **DL 2.1: To what extent does the introduction of a custom loss function improve on a baseline state-of-the-art segmentation solution?**

The development and implementation of a custom loss function based on segmentation topology was shown to positively improve on the SotA baseline by at least 2 – 4% DSC, across the full range of testing data defined in section 3.3. It is also noted that the value of such a loss function could extend outside of this particular domain, where its principle target is to encourage good region separation in tasks involving high density, single-class, elliptical segmentations. An example for this could be blood-cell segmentation.

An important caveat to the work on this function is that the implementation described by this thesis was non-optimised, and therefore additional computational costs should be considered when considering the use of such a function, in its current form.

- **DL 2.2: To what extent does the introduction of a custom loss function support the generation of a 3D representation of unique corallites?**

The topological metric set out in this thesis was shown to be ineffective for describing the small, but meaningful changes produced by the topological loss function, particularly as accuracy approached 1. However, qualitative evaluation showed a positive improvement in corallite region separation and discussions throughout the implementation and evaluation of the corallite reconstruction phase emphasised the importance of this separation for accurate model generation within a 3D environment. Additional metrics were proposed, and surface-level experimentation suggested positive improvements to both total corallite count error and geometric compactness. It is noted that future work might look at validating these other metrics further, to provide a better quantitative measure of the topological accuracy this loss produces.

The exploration and implementation of this second phase of the project identified a number of exciting research opportunities, and introduced some new and intriguing questions. These are noted in the future opportunities summary in section 5.2, but suggest that there is plenty of scope for developing the work in this thesis and continuing towards a comprehensive spatial-temporal solution to corallite reconstruction.

The third, and final phase of this project considered the automated reconstruction of individual corallites, given a segmentation volume of region predictions. The implementation of one possible approach was discussed in section 3.3, and evaluated in detail in section 4.3. The prototype approach showed that individual corallite reconstruction was indeed viable. The implementation proposed is parameteriseable at various stages of the reconstruction process, and this work recommends further input from corallite experts to better define the bounds of these parameters, and ensure as much model accuracy as possible. However, the branching-merging behaviours were shown to be well captured by the implementation, and over 100,000 unique corallites from the Naturalis 6781 colony were modelled with complete freedom of movement and corallite isolation capabilities.

5.2 Future Opportunities

The work in this thesis selected a number of key areas of interest within the bounds of the time available to it. In doing so, a range of areas which offer interesting opportunities for future work have been identified and will be discussed below. The opportunities are presented in no particular order, but under the two broad categories of deep-learning and corallite reconstruction, in reference to the latter two phases of the project pipeline.

5.2.1 Deep-Learning Opportunities

The follow set of points relate directly to the deep-learning phase of this project, including the data, and areas of interest for deeper exploration.

Dataset Expansion: This project introduced a novel set of data, which has been shown to have significant effect on improving a full coverage segmentation of all corallite regions in the μ CT data. Due to time constraints, these were established in the context of a minimum viable dataset to achieve the goals of this thesis. As discussed in section 4.1, additional annotations would comprise an ideal minimum of at least two, non-adjacent, complete slices from each projection angle of a given colony. Future work might look to verify that this minimum training set is sufficient to reproduce all individual corallites of any colony of the *Porites* genus, and further consider if this property extends to colonies of other genera in the Scleractinia order.

TCM Study: The depth of the volumetric snippets (3.1.3) used in this thesis was led by work in [42], however the domain is quite different. Where the original implementation considered frames in videos, this project considers the volume spatially, at μm scale. It is very plausible that important contextual information extends beyond the depth of 5 slices. A full study of the performance implications of this parameter would provide valuable insights into the extent of volumetric corallite relationships.

ViT Input/Patch Size: Work in [5] suggests that an input size of 512×512 and a patch size of 8 gave the best results for their task. It was noted that the additional computational trade-off was high, for little improvement, but the work suggests exploration of these parameters might offer further performance gains.

Topological Loss Optimisation: The implementation of the topological loss calculation has not been optimised for computational performance, particularly for GPU architectures. The additional computational demands caused an increase in training time by a factor of at least 16, exceeding 4 days per training schedule for the VLrg models using topological loss. This had a significant effect on the rate at which experiments could be conducted and evaluated, and thus inhibited the depth at which some topics could ideally be explored. Proper optimisation of the loss calculations would provide opportunity for more thorough research on its application.

Topological Metric Review: The topological metric defined in section 2.2.4 was shown to be unsuitable for measuring fine-grained changes in overall topological accuracy. Development of a metric which better describes the value of marginal improvements, once a good baseline topology is achieved, would be invaluable in validating future work in this domain.

Data and Topological Loss Relationships: The main focus of experimentation with respect to the topological loss implementation was to understand the effects of the weighting applied to the loss, ie. the value of λ . Evaluation of these experiments highlighted two interesting potential correlations; One between the effect of the topological loss, and the amount of overlap in the training data. The other between the effect of the topological loss and the length of the warm-up period. There is interesting future study in validating these correlations, and if they exist, understanding the nature of these relationships with respect to developing similar models in this domain.

Gaussian Noise Testing: This thesis considered one other species of the *Porites* genus as testing data to understand how well the model might generalise across species. The addition of Gaussian noise was shown to improve generalisation well, however it remains an open question as to whether this improvement is consistent across other species of *Porites* corals.

5.2.2 Corallite Reconstruction Opportunities

The following set of points relate directly to the corallite reconstruction phase of the thesis, ie. the automated generation of a three-dimensional model, given a volume of corallite region predictions via the deep-learning pipeline discussed throughout this thesis, or otherwise.

Optimisation of Corallite Tracing Process: The corallite tracing process described in section 4.3 has not been optimised, and as a result, takes significant time (hrs) to form a complete mapping of segmentations to unique corallites. There is certainly scope to optimise this process, particularly when searching for the most likely region connection in the previous layer. A simple first step would be to reduce the search in slice $n - 1$ to a small radius from the center of the region in slice n to which a corallite mapping is sought. Naturally, some time spent on this optimisation would uncover a number of interesting algorithms which might additionally be employed.

Parameterisation of Corallite Tracing: Exploration of the parameters used to associate regions in the prediction volume to with unique corallites, as discussed in section 4.3, would help to define suitable constraints to the corallite tracing process. Ideally, this would be explored with guidance from an expert in corallite development. Additional properties of the segmentation regions could be leveraged to improve on and build a more complex model of the development of corallite structures;

- Comparisons of the orientation of the ellipse mapped to regions, as the direction of real corallite growth should be consistent within some bounds.
- Considerations of the k-nearest neighbours in slice $n - 1$ might help to better identify when corallites split and merge, thus building a more effective model of the branching behaviours observed.

Deep-Learning for Corallite Tracing: An alternative approach could be to explore a secondary layer of deep-learning, to train a model to identify and assign unique corallite id's, given a segmentation volume. This would require extensive annotation work on some segmentation outputs, in order to accurately describe the variety of corallite paths observed in the scans, but has the potential to model the nuances of the corallite pathways more effectively than the algorithmic approach taken by this thesis.

NURBS-Based Model Generation: Discussion in section 4.3 noted that computational efficiency was potentially a significant barrier to the pipeline. Alternate forms of geometric representation might be explored, such as the use of NURBS, to model corallite growth. This work could consider evaluating the differences between the two approaches in terms of model complexity, and computational efficiency.

Full Spatial-temporal Reconstruction: Complete spatial-temporal reconstruction is, perhaps, the ultimate goal of a system such as this. In order to achieve this goal, a secondary learning pipeline would need to be developed. This pipeline would learn to determine the rate of growth, by the density of the corallite structure along the projection axis. Conceptually, if the rate of growth is high in a given year, the density is lower, and vice-versa if the rate of growth is low. By learning these properties, markers can be integrated into the model of each corallite describing the amount of growth in a given period of time. This system would provide a clearer model of not only where each corallite has grown over time, but also at what rate.

Bibliography

- [1] Kenneth R. N. Anthony, Sean R. Connolly, and Bette L. Willis. Comparative analysis of energy allocation to tissue and skeletal growth in corals. *Limnology and Oceanography*, 47(5):1417–1429, 2002. [doi:
https://doi.org/10.4319/lo.2002.47.5.1417](https://doi.org/10.4319/lo.2002.47.5.1417).
- [2] Maria Carla Benedetti, Lorenzo Bramanti, Cristina Priori, Fabrizio Erra, Mimmo Iannelli, Fabio Bulleri, and Giovanni Santangelo. Polyp longevity in a precious gorgonian coral: hints toward a demographic approach to polyp dynamics. *Coral Reefs*, 39:1125–1136, 2020.
- [3] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.
- [4] Herman Cesar, Lauretta Burke, and Lida Pet-Soede. The economics of worldwide coral reef degradation. 2003.
- [5] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.
- [6] WM Darke and DJ Barnes. Growth trajectories of corallites and ages of polyps in massive colonies of reef-building corals of the genus porites. *Marine Biology*, 117:321–326, 1993.
- [7] Elizabeth Dávalos-Dehullu, Héctor Hernández-Arana, and Juan P Carricart-Ganivet. On the causes of density banding in skeletons of corals of the genus montastraea. *Journal of experimental marine biology and ecology*, 365(2):142–147, 2008.
- [8] Jia Deng, R. Socher, Li Fei-Fei, Wei Dong, Kai Li, and Li-Jia Li. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, volume 00, pages 248–255, 06 2009. URL: <https://ieeexplore.ieee.org/abstract/document/5206848/>, doi:10.1109/CVPR.2009.5206848.
- [9] Li Deng, Dong Yu, et al. Deep learning: methods and applications. *Foundations and trends in signal processing*, 7(3–4):197–387, 2014.
- [10] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [11] Omar Elharrouss, Younes Akbari, Noor Almaadeed, and Somaya Al-Maadeed. Backbones-review: Feature extraction networks for deep learning and deep reinforcement learning approaches. *arXiv preprint arXiv:2206.08016*, 2022.
- [12] Paul Fergus and Carl Chalmers. Image classification and object detection. In *Applied Deep Learning: Tools, Techniques, and Implementation*, pages 173–201. Springer, 2022.
- [13] Greg D Field and EJ Chichilnisky. Information processing in the primate retina: circuitry and coding. *Annu. Rev. Neurosci.*, 30:1–30, 2007.
- [14] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980.
- [15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 630–645, Cham, 2016. Springer International Publishing.
- [16] Ove Hoegh-Guldberg, Elvira S Poloczanska, William Skirving, and Sophie Dove. Coral reef ecosystems under climate change and ocean acidification. *Frontiers in Marine Science*, 4:158, 2017.

- [17] N. Kanopoulos, N. Vasanthavada, and R.L. Baker. Design of an image edge detection filter using the sobel operator. *IEEE Journal of Solid-State Circuits*, 23(2):358–367, 1988. [doi:10.1109/4.996](https://doi.org/10.1109/4.996).
- [18] Salman Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan, and Mubarak Shah. Transformers in vision: A survey. *ACM Comput. Surv.*, 54(10s), sep 2022. [doi:10.1145/3505244](https://doi.org/10.1145/3505244).
- [19] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL: <http://arxiv.org/abs/1412.6980>.
- [20] Marcelo V Kitahara, Stephen D Cairns, Jarosław Stolarski, David Blair, and David J Miller. A comprehensive phylogenetic analysis of the scleractinia (cnidaria, anthozoa) based on mitochondrial co1 sequence data. *PloS one*, 5(7):e11490, 2010.
- [21] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. [doi:10.1109/5.726791](https://doi.org/10.1109/5.726791).
- [22] Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer, and Tom Goldstein. Visualizing the loss landscape of neural nets. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL: https://proceedings.neurips.cc/paper_files/paper/2018/file/a41b3bb3e6b050b6c9067c67f663b915-Paper.pdf.
- [23] Yixin Li, Tingyu Han, Kun Bi, Kun Liang, Junyuan Chen, Jing Lu, Chunpeng He, and Zuhong Lu. The 3d reconstruction of pocillopora colony sheds light on the growth pattern of this reef-building coral. *Iscience*, 23(6):101069, 2020.
- [24] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943. [doi:10.1007/BF02478259](https://doi.org/10.1007/BF02478259).
- [25] Francisco Medellín-Maldonado, Andrés López-Pérez, Leopoldo Ruiz-Huerta, and Juan P. Carricart-Ganivet. Understanding corallite demography to comprehend potential bias in sclerochronology: Analysis of coral modular growth by micro-computed tomography. *Limnology and Oceanography*, 12 2022. [doi:10.1002/lno.12229](https://doi.org/10.1002/lno.12229).
- [26] National Oceanic and Atmospheric Administration. Coral reef ecosystems, 2019. URL: <https://www.noaa.gov/education/resource-collections/marine-life/coral-reef-ecosystems>.
- [27] David Ojika, Bhavesh Patel, G. Anthony Reina, Trent Boyer, Chad Martin, and Prashant Shah. Addressing the memory bottleneck in AI model training. *CoRR*, abs/2003.08732, 2020. URL: <https://arxiv.org/abs/2003.08732>, arXiv:2003.08732.
- [28] Michel Pichon. *Porites*, pages 815–821. Springer Netherlands, Dordrecht, 2011. [doi:10.1007/978-90-481-2639-2_273](https://doi.org/10.1007/978-90-481-2639-2_273).
- [29] Gopinath Rebala, Ajay Ravi, and Sanjay Churiwala. *Machine Learning Definition and Basics*, pages 1–17. Springer International Publishing, Cham, 2019. [doi:10.1007/978-3-030-15729-6_1](https://doi.org/10.1007/978-3-030-15729-6_1).
- [30] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *CoRR*, abs/1804.02767, 2018. URL: <https://arxiv.org/abs/1804.02767>, arXiv:1804.02767.
- [31] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015. URL: [http://arxiv.org/abs/1505.04597](https://arxiv.org/abs/1505.04597), arXiv:1505.04597.
- [32] Ainsley Rutherford, Leonardo Bertini, Erica J. Hendy, Kenneth G. Johnson, Rebecca Summerfield, and Tilo Burghardt. Towards the analysis of coral skeletal density-banding using deep learning. *SN Applied Sciences*, 4(2):38, 2022. [doi:10.1007/s42452-021-04912-x](https://doi.org/10.1007/s42452-021-04912-x).
- [33] Dominik Scherer, Andreas Müller, and Sven Behnke. Evaluation of pooling operations in convolutional architectures for object recognition. In *Artificial Neural Networks–ICANN 2010: 20th International Conference, Thessaloniki, Greece, September 15–18, 2010, Proceedings, Part III* 20, pages 92–101. Springer, 2010.

BIBLIOGRAPHY

- [34] Connor et al. Schmidt. Faster crystallization during coral skeleton formation correlates with resilience to ocean acidification. *Journal of the American Chemical Society*, 2022. [doi:10.1021/jacs.1c11434](https://doi.org/10.1021/jacs.1c11434).
- [35] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. [doi:10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x).
- [36] Adam Smith, Nathan Cook, Kailash Cook, Rachelle Brown, Richard Woodgett, John Veron, and Vicki Saylor. Field measurements of a massive porites coral at goolboodi (orpheus island), great barrier reef. *Scientific Reports*, 11(1):15334, 2021. [doi:10.1038/s41598-021-94818-w](https://doi.org/10.1038/s41598-021-94818-w).
- [37] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [38] Daniel Wagner, Alan M. Friedlander, Richard L. Pyle, Cassandra M. Brooks, Kristina M. Gjerde, and T. ‘Aulani Wilhelm. Coral reefs of the high seas: Hidden biodiversity hotspots in need of protection. *Frontiers in Marine Science*, 7, 2020. URL: <https://www.frontiersin.org/articles/10.3389/fmars.2020.567428>, doi:10.3389/fmars.2020.567428.
- [39] Paul J. Werbos. Backpropagation: past and future. In *Proceedings of International Conference on Neural Networks (ICNN’88), San Diego, CA, USA, July 24-27, 1988*, pages 343–353. IEEE, 1988. [doi:10.1109/ICNN.1988.23866](https://doi.org/10.1109/ICNN.1988.23866).
- [40] D.R. Wilson and T.R. Martinez. The need for small learning rates on large problems. In *IJCNN’01. International Joint Conference on Neural Networks. Proceedings (Cat. No.01CH37222)*, volume 1, pages 115–119 vol.1, 2001. [doi:10.1109/IJCNN.2001.939002](https://doi.org/10.1109/IJCNN.2001.939002).
- [41] Yanzhao Wu, Ling Liu, Juhyun Bae, Ka Ho Chow, Arun Iyengar, Calton Pu, Wenqi Wei, Lei Yu, and Qi Zhang. Demystifying learning rate polices for high accuracy training of deep neural networks. *CoRR*, abs/1908.06477, 2019. URL: <http://arxiv.org/abs/1908.06477>, arXiv:1908.06477.
- [42] Chengxi Zeng, Xinyu Yang, Majid Mirmehdi, Alberto M Gambaruto, and Tilo Burghardt. Videotransunet: Temporally blended vision transformer for ct vfss instance segmentation. *arXiv preprint arXiv:2208.08315*, 2022.