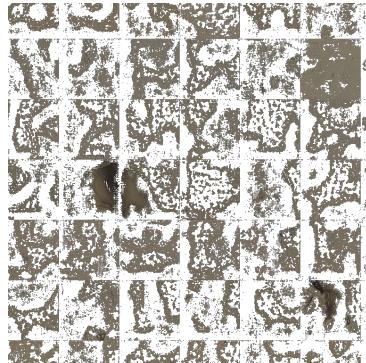


QCES Part III Project: Using deep learning for processing high-resolution sediment core images from the Iberian Margin



Sam Hutton

Due: 18/05/2025
Word Count: 9978

Abstract

Sediment cores contain a long-timescale record of important climatological parameters. Even common, expensive, time-consuming sampling methods cannot match the resolution of colour data taken from digital photographs; but this data is noisy, in part due to bioturbation, drilling disturbance, and sampling artifacts. This report presents a number of improvements to existing neural network based methods of segmenting ‘undisturbed’ sediment from core images and applies the newly created model to producing a filtered record of lightness over 160m \sim 1400ka. These improvements include a novel morphological-process background removal, selection of an improved model architecture (SegNet), and three methods for synthetic data augmentation. Data augmentation did not have a significant effect on model performance. With all of these improvements, our new method produces an Intersection-Over-Union of 0.53 ± 0.09 . This model likely underpredicts disturbance, despite selecting 26.6 % of the average core. This suggests a need for more training on a larger dataset to improve the model, and a need for the development of stronger statistical tools to address the gaps created by disturbance modifying the record. Cross-correlation analysis suggests that the resolution available in undisturbed sediment may be lower than previous estimates. Tools for using this model and other common stitching and averaging operations on digital photographs are presented in the `coreclean` module for python, developed alongside this report.

Declaration

This report is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any work that has already been submitted, or, is being concurrently submitted for any degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. It does not exceed the prescribed word limit. I also confirm that I have read and understood the Faculty of Mathematics Guidelines on Plagiarism and the University-wide Statement on Plagiarism.

Signed: Sam Hutton. Dated: 2025-05-18.

Contents

1	Introduction	3
1.1	Background	3
1.2	Aims	6
2	Methods	7
2.1	Dataset	7
2.2	Colour Transforms	7
2.3	Preprocessing	8
2.4	Convolutional Neural Networks	10
2.5	Data Augmentation	13
2.6	Training	14
2.7	Post-Processing	14
2.8	Implementation	15
3	Results	16
3.1	Model Architecture	16
3.2	Background Removal	17
3.3	Data Augmentation	17
3.4	Confidence Maps	19
3.5	Filtered Colour Record	19
3.6	Qualitative Comparison	19
4	Discussion	22
4.1	Method Usage & Generality	22
4.2	Model Performance	22
4.3	Model Limitations	23
4.4	Implications for Future Sampling	24
4.5	Implications for Resolution of Colour Records	24
5	Conclusions	26
5.1	Future Work	26
5.2	Final Summary	27
A	Dependencies	32
B	Code	33

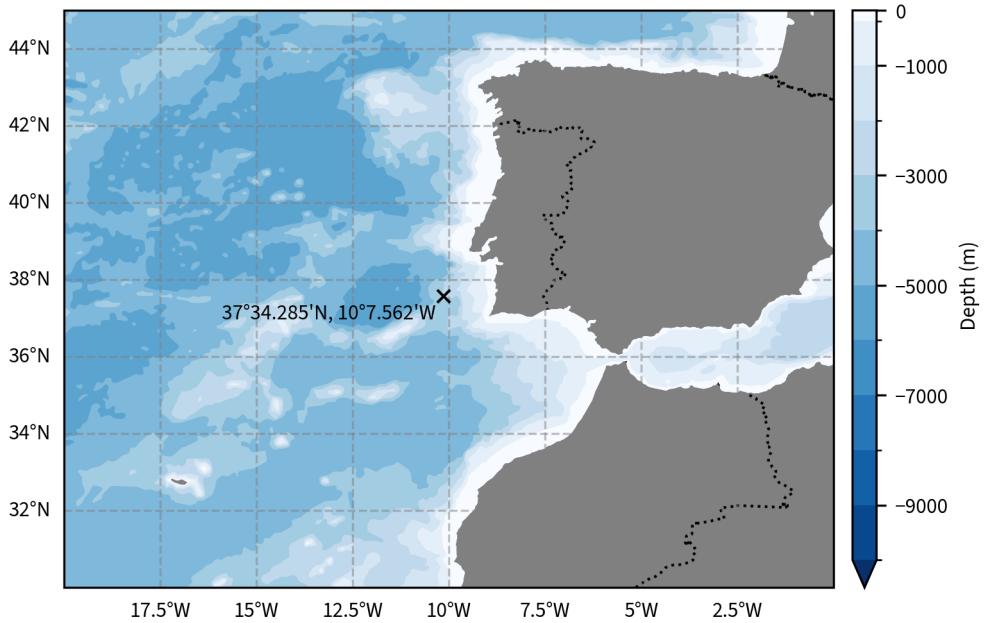


Figure 1: Location of site U1385 on the Iberian Margin.

1 Introduction

Marine sediment cores contain a high-resolution long timescale record of Earth's climate history (Giosan et al., 2002; Zhang et al., 2023). These records, especially those of the Iberian Margin are useful in parametrising abrupt millennial scale events (Hodell et al., 2013; 2015), which provides training data for future modelling. Spectrophotometry is used to obtain colour reflectance data, which displays correlation with dust and carbonate content in this core. A higher resolution of reflectance data is theoretically available from digital photographs of core sections (Chapman & Shackleton, 1998; Nederbragt et al., 2006; Wu et al., 2022). Colour records tend to be noisy, partly due to bioturbation, detritus, drilling disturbances and oxidation, collectively referred to as 'disturbance' throughout this report (Expedition 339 Scientists, 2013a). It is possible to segment this disturbance out of small patches of cores using Convolutional Neural Networks (CNNs), but no effort has yet been made to apply this technique to generating full core records free of bioturbation, nor has it been applied to the Iberian Margin (Fazekas et al., 2017). Using shipboard images from IODP Expedition 399, site U1385 "Shackleton Site", this report aims to train a CNN to segment disturbance and generate an undisturbed record of sediment colour variation. We also apply a number of improvements and optimisations in hopes of producing a better model which can provide a more accurate record.

1.1 Background

1.1.1 Climate Relevance of Colour Data

The colour of sediment cores is useful for a number of paleoclimatological parameters; the lightness (L^*) shows a very high correlation with the log of the Calcium-Titanium ratio ($\log \text{Ca/Ti}$, Giosan et al. (2002)), and has been interpreted as a proxy for the weight percent calcium carbonate (% CaCO_3) in a core (Gebregiorgis et al., 2020). This signal displays a signal tunable to precession which can help in dating. Further, the $\log \text{Ca/Ti}$ mirrors sea surface temperature estimates from alkenones and planktonic $\delta^{18}\text{O}$ (Hodell et al., 2015).

The carbonate signal is thought to be controlled by the input of detrital sediments, i.e. the percentage of carbonate decreases when the flux of clay from riverine, atmospheric and margin slopes increases, despite the biological input of carbonate remaining constant through time (Thomson et al., 1999). It is suggested that the processes that control clay deposition are enhanced by a low sea level, and this in turn is affected by Milankovitch frequencies, so that on orbital timescales, the driver of this variability is the detrital input. On millennial timescales, however, it is likely that productivity affects the carbonate record. During glacial stadials, cold sea surface temperatures (SSTs) are correlated with low bulk carbonate $\delta^{18}\text{O}$

and more reworked nanofossil taxa, suggesting a drop in productivity. ([Hodell et al., 2013; Incarbona et al., 2010](#)).

In either case, lightness and the carbonate record display a strong precession cycle, which has a lag of about 2.8kyr compared to the modelled precession from radiocarbon. This means we can use colour records to tune the timescale of sediment records, as well as providing a reflection of carbonate productivity and therefore SST over stadials. Millennial variability in log Ca/Ti is persistent in this core, ([Hodell et al., 2015](#)) so high resolution colour data may be able to provide more insight into the shape and features of this variability.

Sediment redness (a^*) can also be used as a paleoclimate proxy ([Deaton & Balsam, 1991; Giosan et al., 2002](#)). In this core, there is a precession signal in a^* , thought to be recording aeolian sourced hematite from African dust. [Hodell et al. \(2013\)](#) suggest that minima in precession provoke aridity on the Moroccan coast. Especially when insolation maxima lead to intense heating and low pressure, this causes more dust deposition. This dust deposition is the dominant mechanism in Moroccan cores ([Bozzano et al., 2002](#)), though the increased seasonal variability at precession minima likely release dust availability. There is probably also some redness signal from fluvial transport increases of hematite due to precipitation and erosion; which would be consistent with modelled predictions of increased runoff at precession minima ([Meijer & Tuenter, 2007](#)), though this could also enhance African dust sources by wet deposition mechanisms.

As with lightness, the redness signal also shows a strong in-phase precession signal, without a significant lag, probably through the atmosphere. This is powerful for tuning, despite its attenuation in the lower parts of the U1385 core ([Hodell et al., 2013; 2015](#)). Paleoclimate reconstruction using colour will help us better understand the high resolution of the variability present in these cores, especially as digital cameras should be able to provide higher resolution.

1.1.2 High Resolution Marine Sediment

The highest resolution of climate data currently available in a sediment core tends to be obtained by X-ray Fluorescence (XRF) techniques, for example at 1cm resolution at Shackleton Site ([Hodell et al., 2015](#)). The colour reflectance spectrophotometric measures on this core, for example, are only taken on 2cm intervals on the fresh core ([Expedition 339 Scientists, 2013a](#)). The work of [Chapman & Shackleton \(1998\)](#) suggests that colour data contains information at the 1cm scale, and possibly even lower, from a correlation-based analysis of spectrophotometer series over one core section; this resolution is obscured by disturbance.

Some of the noise of a spectrophotometer is removed by averaging over the large window size; theoretically, a well calibrated digital camera should be able to generate high quality colour data of a higher resolution ([Nederbragt et al., 2006](#)) but obtaining higher resolutions is useless if noise dominates the signal. Reflections from lighting and the presence of disturbance mean that high resolution records from photographs show significant artifacts ([Wu et al., 2022](#)).

1.1.3 Processing Sediment Images

Disturbance can be visually identified in images of sediment cores and has been qualitatively rated for site U1385 ([Expedition 339 Scientists, 2013b](#)). Some work has been done to use digital image software to manually remove bioturbation and quantify ichnological disturbance; this is time consuming and prone to error, and the parameters selected are only valid for individual core sections or even specific areas within cores ([Dorador & Rodríguez-Tovar, 2014; 2016; 2018](#)). Neural networks and deep learning strategies may be able to select bioturbation more generally.

[Fazekas et al. \(2017\)](#) used a Fully Convolutional Network (FCN) architecture to segment drilling disturbance and accessories from patches of linescanner images of cores from site U1308, and obtained a modest Intersection-Over-Union (IOU) score of 30.19%, but showed that the majority of manually identified artifacts were detectable by the FCN in these patches. This was attributed mostly to the model selecting true phenomena as disturbance.

Since the development of FCNs for segmentation by [Long et al. \(2015\)](#), a number of new Convolutional Neural Network (CNN) model architectures have been developed to decrease training time and increase accuracy. Notably, UNet architectures, originally developed for biomedical applications allow the evaluation of larger images in faster time and with less labelled data. ([Ronneberger et al., 2015](#)). Further, SegNet consumes less memory and provides similar accuracy with fewer training epochs ([Badrinarayanan et al., 2015](#)).

Additionally, since this previous work, the practice of data augmentation by applying transforms to training images has become commonplace in machine learning, which allows training on smaller datasets ([Mumuni & Mumuni, 2022](#)). This is particularly useful with difficult to obtain records such as sediment cores; and can help prevent overfitting when working with these smaller datasets. Newer pipelines for biomedical image segmentation often involve preprocessing steps that reduce the amount of data to be processed by more sophisticated and computationally intense techniques, which makes segmentation a less-costly and more time-efficient task ([Neha et al., 2024](#)).

1.1.4 Site U1385

Shackleton site (U1385, Fig 1), located on the Western Iberian Margin ($37^{\circ}34.285'N$; $10^{\circ}7.562'W$) at 2585mbsl was drilled on the 25 November 2011 during IODP expedition 339 ([Expedition 339 Scientists, 2013a](#)). Centennial to millenial scale changes are present in records from this site due to the high sedimentation rate of 10-20cm/ka. Historically, this site has had utility because of the unambiguous link between both hemisphere's ice core records. The upper water column and SST, recorded by planktic foraminifera and alkenones, resembles the ice core records of Greenland, whereas the benthic foraminifera produce $\delta^{18}\text{O}$ signals resembling Antarctic ice records, likely because of the ice volume signal ([Shackleton et al., 2000](#)).

Cores recovered from IODP expeditions are given a numbering, consisting of the site, U1385, followed by the hole letter, core number, and then the section, e.g. U1385A-01H-01. This convention will be used throughout this report.

1.1.5 Core Disturbances

A number of secondary features obscure the primary colour signal in U1385 sediment (Fig. 2). The major features which we labelled as disturbance for the purposes of this method are given below.

1. Drilling disturbance, broadly classified as one of seven types ([Expedition 339 Scientists, 2013a](#)):
 - a) **Fall-in:** Material from the top of the core falls downhole onto the cored surface.
 - b) **Bowed:** Bedding contacts are bent from horizontal due to core barrel pressure.
 - c) **Flow-in:** Coring slurry flows into the core around the undisturbed sediment.
 - d) **Soupy or Mousselike:** Water saturation has destroyed the structure of the sediment.
 - e) **Biscuit:** Sediments of higher stiffness are relatively undisturbed within soupy intervals.
 - f) **Cracked or Fractured:** Firm sediments are broken by the coring process but not displaced.
 - g) **Fragmented or Brecciated:** Sediments are pervasively broken into fragments and displaced.
2. Lithological Accesories:
 - h) **Macrofossils:** Mostly gastropod shell fragments, some sand-sized foraminifera.
 - i) **Nodules and Concretions:** Pyritised burrows and nodules.
3. Shipboard sampling artifacts, incurred by sampling and preparation of the core:
 - j) **Slide Scratches:** Using a microscope slide for cleaning the surface is common, but can cause scrapes in the surface introducing reflections.
 - k) **Headspace samples:** Large chunks of sediment are removed from the bottom of core sections for gas sampling.
 - l) **Cutting scrapes:** When halving the core, hard chunks can be dragged up the core, causing gouges in the sediment.
4. Bioturbation, generally discrete burrows of four ichnotaxa ([Rodríguez-Tovar & Dorador, 2014](#)):
 - m) **Thalassinoides:** Large, mostly horizontal branched burrows, generally observed as 6-11mm subcircular ovals with straight cylinders 22-43mm long. Produced mostly by decapods and other crustaceans.
 - n) **Chondrites:** Secondary burrows, appearing as clusters of roughly circular holes 0.5-3mm in diameter, mostly only in deep sections, usually considered an unknown creature's feeding structure.
 - o) **Zoophycos:** Helical structures, demonstrating horizontal lobes (spreiten structures). Often comes in a series of stacked lamina, and have several interpretations,
 - p) **Planolites:** Smaller (2-7mm) subcircular horizontal burrows, interpreted as produced by soft bodied invertebrates.
5. Further disturbance. There are a number of smaller or less frequently observed artifacts, including smears, water pools, and less common ichnotaxa such as *Phycosiphon*, *Scolica*, *Taenidium* and *Paleophycus*.

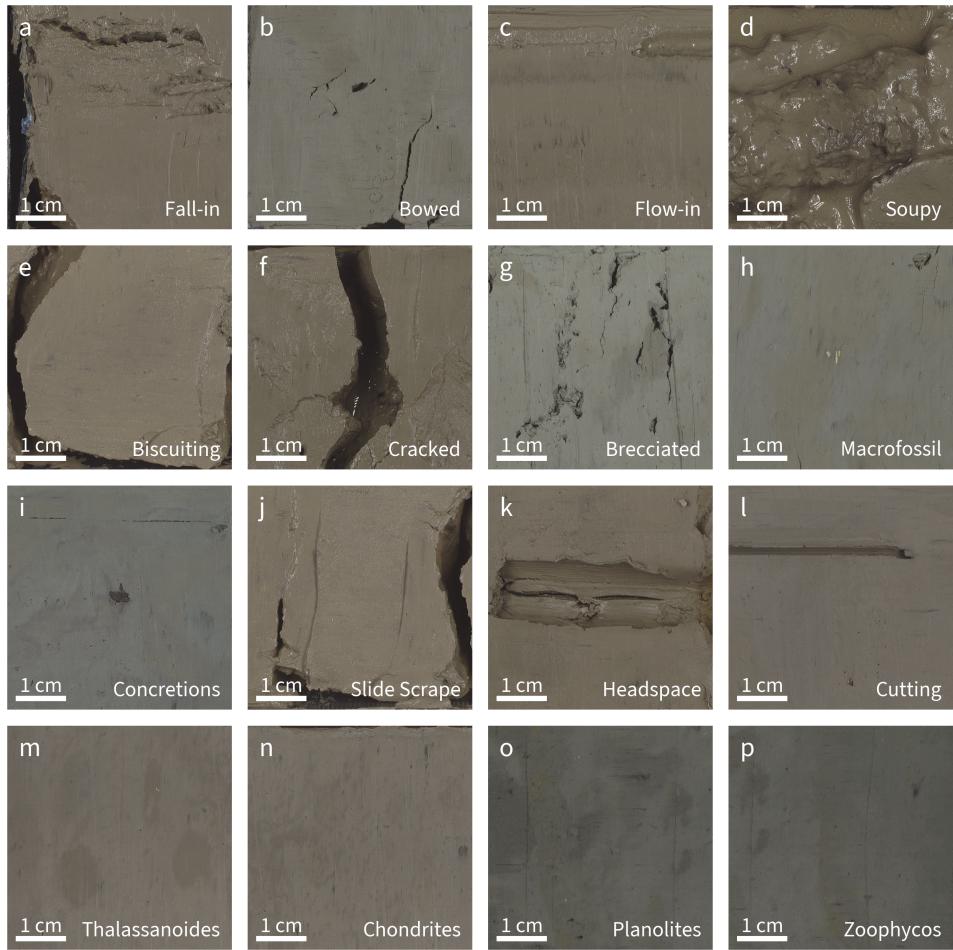


Figure 2: Types of disturbance in U1385 cores. All images are U1385, specifically sections (a) A-02H-1 (b) D-09H-1 (c) B-01H-3 (d) A-03H-1 (e) A-01H-1 (f) A-02H-2 (g) B-13H-7 (h) D-11H-6 (i) E-14H-1 (j) A-01H-1 (k) A-02H-6 (l) A-04H-4 (m) E-04H-4 (n) E-04H-4 (o) E-13H-4 (p) E-13H-4.

1.2 Aims

By developing an improved method for disturbance segmentation, including pre-processing steps, selecting an optimised model architecture, applying synthetic data augmentation and then post-processing the obtained images into a colour record, this report aims to create a new method for processing core images to high resolution colour data, training on IODP Expedition 339's U1385 cores. We aim to optimise and improve the segmentation performance of CNNs for the task of marine sediment core segmentation by evaluating various novel improvements including algorithmic morphological background removal, pre-process thresholding, and three strategies of data augmentation. We will explore the resolution that is possible to obtain from photo data, and discuss the implications for future use of sediment colour for paleoclimate.

2 Methods

2.1 Dataset

The primary dataset used for this work was the shipboard images of the Expedition 339 cores from Shackleton sites. The cores in our dataset were extracted using the advanced piston corer (APC), which provides less coring disturbance than the previously used extended core barrel (XCB) system when used on soft sediment. Images were taken using the Section Half Imaging Logger (SHIL), at a resolution of 20 lines at 500 dpi, providing us with a three channel (RGB) image in TIFF format ([Expedition 339 Scientists, 2013a](#)). Including core-catcher segments (small core sections from the end of drilling a hole), this generated 503 images. The 65 core catcher images are approximately 25cm each; the remaining 438 are approximately 1.2-1.5m each, reaching a maximum depth of 162m below sea-level once corrected, with an age estimated at 1422ka ([Hodell et al., 2015](#)).

Of these images, 12 were labelled manually and in detail to remove disturbance using GNU Image Manipulation Program (GIMP). The alpha layer was used to contain the image mask ([Fazekas et al., 2017](#)). There were four main passes in this labelling step. The first involved using a rectangle select tool to select and mask the background and edges of the core; second, obvious drilling disturbance, oxidation, surface disturbances, voids and other clear artifacts were removed. For identifying bioturbation, following [Dorador & Rodríguez-Tovar \(2018\)](#), the colour of the image was modified using the levels tool, and trace fossils were labelled as disturbance. Finally, the newly generated labels were referenced with the original visual descriptions; in locations where the bioturbation index and disturbance from the visual descriptions did not match our labels, our labels were updated ([Expedition 339 Scientists, 2013b](#)). Labelling was a time-consuming step, and the labels are not perfect; especially, in sections where bioturbation was questionable, no labels were applied to preserve the maximum amount of data.

All of the labelled cores from the composite section were excluded from the training set to avoid overfitting. These were cut into patches of 128x128 (giving 27,222 patches or 18,239 after background removal) and 256x256 (giving 4364 patches after background removal). These patch sizes were selected because square images are easier to process with CNNs and 128 and 256 are commonly used sizes in segmentation tasks ([Fazekas et al., 2017; Ronneberger et al., 2015](#)). Based on the work of [Fazekas et al. \(2017\)](#), most models were trained using the 256x256 images, as this gives us patches of approximately 1.2cm, roughly the scale of most disturbance; using the 128x128 patches produced overfitting and artifacting (Sec. [4.3.2](#)). This dataset size is comparable to, but smaller than the commonly used Oxford IIIT Pets dataset, which has \approx 7400 images ([Parkhi et al., 2012](#)), though the quality of the labelled data is likely lower, and there is more correlation between images, as they are taken from larger original core sections.

2.2 Colour Transforms

Our data was provided in .tif format, with colour in a linear RGB space. RGB images measure the intensity of red, green and blue light from 0 to 255, in a space originally designed to match the power response of CRT displays to voltage changes; it measures the power to emit. This is not ideal for quantifying colour changes as it lacks the property that linear distance in the space appears uniform; for this reason much previous work in colour records uses the CIELAB space; measuring colour in lightness (L^*), and two chromaticity parameters (a^* , redness and b^* , yellowness) which capture the hue of the pixel.

In order to transform linear RGB in our tiff images to CIELAB, we use Eqs. [1 - 5](#), as implemented in the openCV library ([Bradski, 2000](#)) using the XYZ colour space as an intermediate:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} L^* \\ a^* \\ b^* \end{bmatrix} = \begin{bmatrix} 116 \cdot f(Y) - 16 + \delta \\ 500 \cdot (f(X) - f(Y)) + \delta \\ 200 \cdot (f(Y) - f(Z)) + \delta \end{bmatrix} \quad \text{for } Y > 0.008856 \quad (2)$$

$$\begin{bmatrix} L^* \\ a^* \\ b^* \end{bmatrix} = \begin{bmatrix} 903.3 \cdot Y \\ 500 \cdot (f(X) - f(Y)) + \delta \\ 200 \cdot (f(Y) - f(Z)) + \delta \end{bmatrix} \quad \text{for } Y \leq 0.008856 \quad (3)$$

$$f(t) = \begin{cases} t^{1/3} & \text{if } t > 0.008856 \\ 7.787t + \frac{16}{116} & \text{if } t \leq 0.008856 \end{cases} \quad (4)$$

$$\delta = \begin{cases} 128 & \text{for 8-bit} \\ 0 & \text{for floats} \end{cases} \quad (5)$$

2.3 Preprocessing

2.3.1 Background Removal

Traditionally, and for our dataset, pixel locations have been used for background removal to generate a clean cropped image (“LIMS Reports”, 2012). This is accurate enough for visual description of a single set of core images. In this report, to accurately remove the background without including any of the apparatus used to hold the core, we found that this method left voids on either side within the core holder, and was not robust to varying sizes of cores, which means it cannot be used simply for the last section of each core and the ‘core-catcher’.

Therefore, we used Otsu’s method for asymmetric classes with multi-class improvement (Cai et al., 2014; Kittler & Illingworth, 1986; Otsu, 1979) to identify colour clusters, and then performed a morphological opening (Haralick, 1992) on a generated binary mask of ‘core-like’ colours, selecting the largest rectangle. This method worked automatically in 97.6% of the core images with optimal parameters, failing mostly on images where mud exceeded the top section of the core and so merged with colour card. Other failure modes included overcropping on large cracks within the core meaning that the lower end of the core was discarded.

A morphological opening consists of two stages, erosion and dilation. During erosion, a small shape (kernel) of a given size is translated pixel by pixel across a large binary image. If all pixels under the kernel are a 1, then the origin of the kernel (the center) is set to a 1. Dilation is similar, but applies if any of the pixels under the kernel are 1. This means that erosion cuts away at shapes, whereas dilation causes them to expand. Doing both in succession causes small features to disappear, as erosion will remove small shapes completely, and then dilation will restore the larger shapes. Using a square kernel emphasizes straight lines, ideal for our application where the desired output is a large rectangle (Haralick, 1992).

Parameters for this method were determined by running an ensemble of methods (Tab. 1). Additional dilation iterations were found to increase the percentage of the sections with the correct length under this method; this reduces the number of large cracks unfilled, which are hard to segment from the background, and therefore result in a partial crop of the core section.

In evaluating this ensemble, an automated procedure was run to detect background removal failure. The colour card was detected by checking 20% of the leftmost column of the image is white. Overcropping (discarding too much of the core) was detected by comparing with the normal section lengths of 120cm and 150cm (to within 5%, as the sections are not all the same length) and then manual verification of the flagged images. All core-catcher images and final sections of each core were manually verified.

Erosion		Dilation				Failure Mode	
Kernel	Iterations	Kernel	Iterations	Cores Failed	Accuracy (%)	Undercropped	Overscropped
5x5	2	5x5	2	32	93.6	0	32
5x5	3	5x5	3	41	91.8	0	41
5x5	3	5x5	5	13	97.4	0	13
5x5	3	5x5	7	12	97.6	6	6
5x5	3	5x5	8	12	97.6	12	0
5x5	3	5x5	10	36	92.8	36	0
5x5	10	5x5	10	133	73.5	0	133
10x10	3	10x10	3	69	86.2	0	69

Table 1: Optimal background removal parameters when tested over the full set of 503 core sections.

This method is applicable to different sizes of cores, and various backgrounds and positioning. The main constraint is the difference in colours of the background to the core; in this case, we assumed that



Figure 3: Background removal on a representative core section (339_U1385A_01H_01). Erosion was applied with a 5x5 square kernel and 3 iterations, and dilation with a 5x5 square kernel and 8 iterations. 10% tolerance was applied cross-core in the bounding rectangle step.

the core lies in the middle of the lightness distribution (i.e. there are two other classes of lightness, which surround the core), but this is configurable for various backgrounds.

Selection of largest inscribed rectangle within the generated contour was trialed by the algorithm of Marzeh et al. (2019), but this was overly computationally intensive; instead adding a 10% extra crop in the cross-core direction was found to quickly convert a bounding rectangle into a clean crop. Extra cropping was not added in the down-core direction as this would add errors to depth from top of the core. Pixel scaling was used to convert to length of cores. Significant gaps in the down-core direction were not produced by this method. Overall steps taken in the background removal steps are shown in Fig. 3.

In the rest of this work, background removal used the parameters of a 5x5 square kernel with 3 iterations of erosion and 7 iterations of dilation. Automatic detection of failures allowed for manual intervention to correctly crop all images.

2.3.2 Thresholding

With the aim of reducing the use of machine learning parameters to encode simple features, we used a simple thresholding method as used in (Blarr et al., 2024), which selects a class of low lightness values. This method is justified by the prevalence of low lightness secondary features, specifically oxidation in our dataset mostly manifests as dark patches, there are dark lithographic accessories in some sections, bioturbation is sometimes associated with organic rich dark regions, and most notably, cracks make dark shadows in the core image which can easily be removed. It is particularly useful in cracked cores such as U1385-A-01H-01

The threshold was determined by Otsu's method (Otsu, 1979), as we intend only to remove the darkest regions from the core images. Otsu's method generates a threshold that separates two classes with different means. Due to the patching of the dataset, applying an extra mask does not actually reduce the required encoding in the model further than cropping. When comparing thresholded images to our labels, they generated an Intersection-Over-Union (IOU) of 0.20 and an accuracy of 82.15%. This is

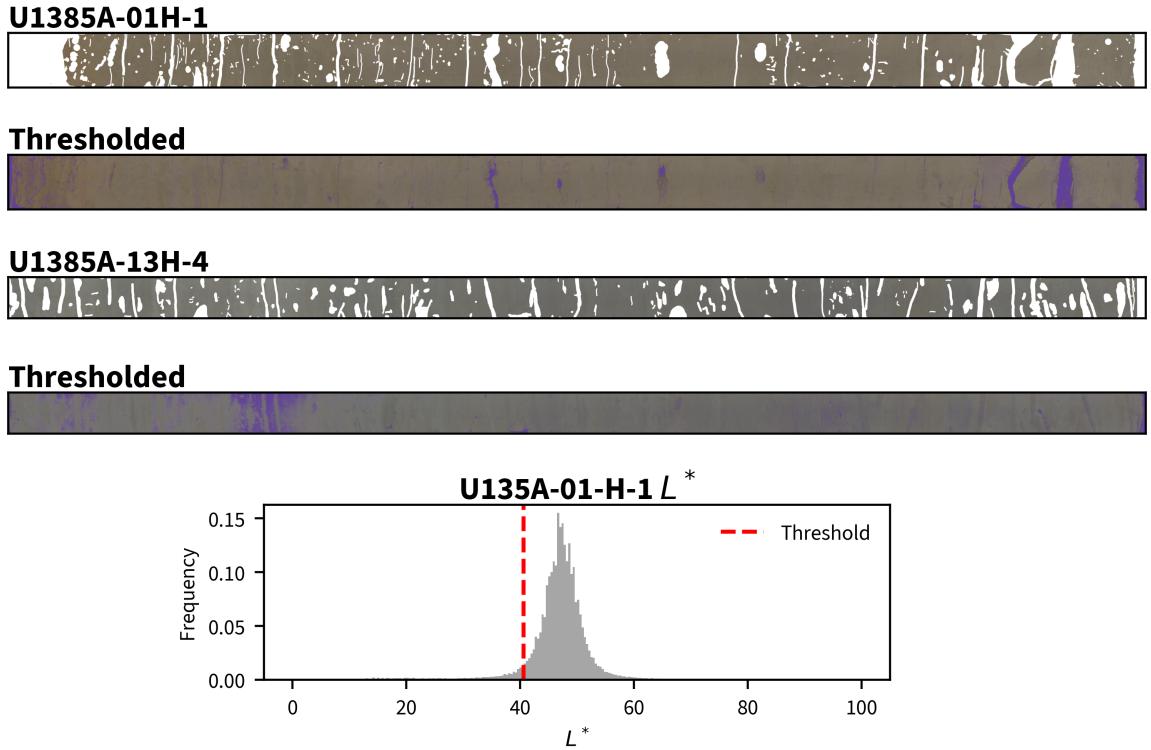


Figure 4: Thresholding applied to two core sections. The histogram shows the distribution of lightness and where the Otsu method places the threshold.

a negligible increase compared to the percentage of all pixels that are undisturbed (i.e. the accuracy of predicting the whole image as undisturbed), 81%. This was not applied to the final model and colour record generation, as in some sections it discarded too much valid sediment. A thresholded image is presented for comparison in Fig. 4.

In some cases, thresholding is useful, especially in heavily cracked core sections such as U1385A-01H-01, but in sections where the majority of the disturbance is bioturbation, such as U1385A-13H-4, the selection of darker sections is quite arbitrary, and there are no significantly lighter classes which can be selected with Otsu's method reliably.

2.4 Convolutional Neural Networks

The convolutional neural net (CNN) framework is well established for image segmentation tasks in biomedical, video processing, and other tasks (Zhao et al., 2024). Essentially, convolutions slide a small patch of weights (a kernel) across a full image, taking the dot product at every step and putting the result into a new 'image'. By applying several different kernels and storing the results in different channels, this can take spatial information and convert it into an encoding with smaller spatial dimension but greater depth (compared to the original 3 RGB channels). This process is repeated several times, alongside pooling which downsamples the image resolution. (Long et al., 2015)

Re-sampling upwards then produces a mask for the image in a matrix with K channels of the same size as the original. Each of those channels represents the confidence (logits) of the model that the image belongs to a certain class. By training, we iteratively update the transformations performed in each convolutional layer to minimise a given metric. Taking a differentiable metric such as Cross Entropy Loss, we can follow gradient descent in our parameter space to slowly tweak the values in each convolutional kernel, until the network gives results close to our desirable output.

CNNs have severable desirable properties; the first is an indifference to the location of features within the patched image, as the same kernel is used across the full patch. Secondly, CNNs are forced to efficiently encode their input, which means that they can incorporate large scale structure, but their convolution steps mean that small scale detail can also be preserved.

Architectures are presented in standard diagrams (Fig. 5,6,7, Iqbal (2018)). Colour coding is used

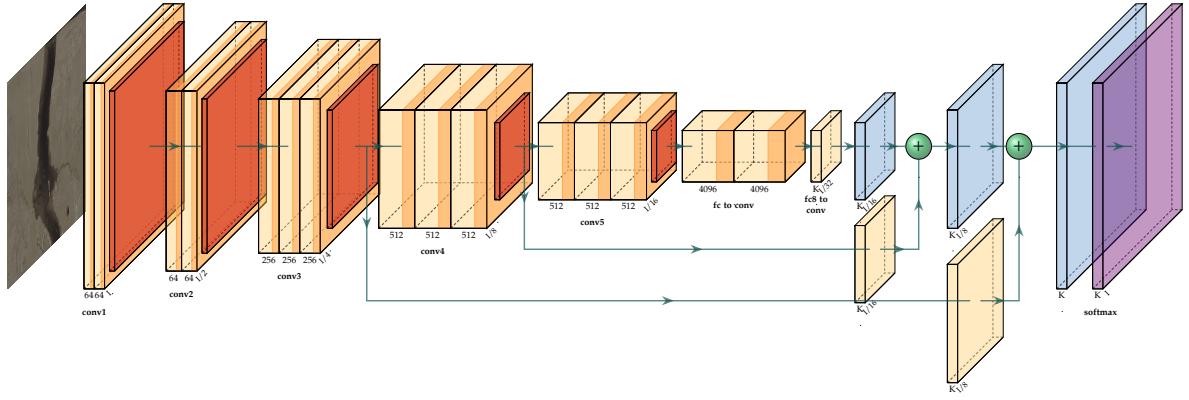


Figure 5: The FCN-8 architecture. ([Iqbal, 2018](#); [Long et al., 2015](#)).

to indicate layer type. As well as the convolution layers, indicated in light orange, max pool layers are indicated in dark orange. These layers downsample the convolved results to reduce the spatial information, but preserve the channel depth created by multi-channel convolution. Blue layers represent upsampling, conducted by choosing nearest neighbours in the low resolution grid. Arrows with right angles represent concatenation, where full encodings from earlier layers are passed to later ones, whereas curved arrows represent passing indices. Finally, purple represents the activation function, converting the arbitrary numbers produced by these operations into a probability of one class or another. The down-sampling stages of a model are known as the ‘encoder’, and the up-sampling sections, the ‘decoder’.

Several model architectures were used in this work and compared.

2.4.1 FCN-8

[Fazekas et al. \(2017\)](#) used the seminal model developed by [Long et al. \(2015\)](#), the Fully Convolutional Model (FCN-8, Fig. 5). This model was a pioneer in convolutional neural networks, as it demonstrated empirically that there was no need for fully connected layers as in previous work, which had simple weights between layers. This intuition was developed by the fact that in most segmentation tasks, it does not matter where in the image the pattern appears, it still should be selected. It notably includes two skips, which allow it to concatenate early embeddings from the convolution of a high resolution image with late data from the pooling.

It essentially takes the backbone of the pre-existing VGG16 ([Simonyan & Zisserman, 2015](#)) model and then applies pooling and upsampling including previous skips to generate a full size model. There is no convolution in the upsampling half of the model, which means that the model is not able to perform spatial agnostic modification on the post-bottleneck encoding of the image. The pre-training makes the fine-tuning of the model slower, as it has a prior understanding of edges and shapes that is not specific to our dataset.

On small datasets, it takes a large number of epochs to converge to a solution, as the number of down-convolution steps have many trainable parameters. ([Badrinarayanan et al., 2015](#); [Yang, 2024](#)).

2.4.2 U-Net

[Ronneberger et al. \(2015\)](#) developed the U-net architecture for biomedical segmentation tasks. It improves on the FCN by replacing the simple pooling in the upsampling half of the network with convolutional layers. Additionally, rather than only 2 skips, the U-net concatenates every layers encoding into the corresponding upsampling layer.

It has improved performance on small datasets, as it can encode more complex features in the upsampling. It tends to be more computationally expensive, and has faced some criticism for being unable to generalise to many types of images. ([Neha et al., 2024](#)). Due to the constraints of this report, the overlap tile strategy proposed by [Ronneberger et al. \(2015\)](#) was not used, rather, just the convolutional architecture. Th tiling method may have allowed the convolution to generalise to larger features, especially those across patches (Sec. 5.1.3).

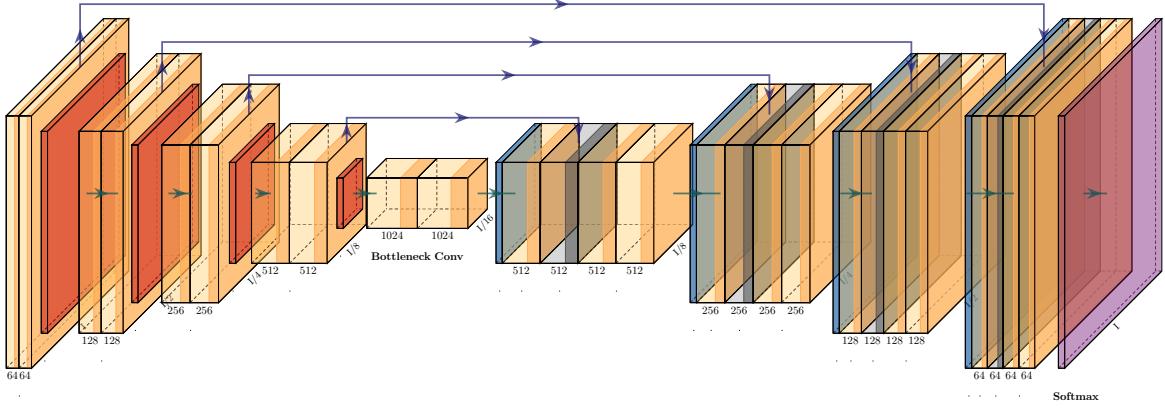


Figure 6: U-Net 32 architecture. ([Iqbal, 2018](#); [Ronneberger et al., 2015](#)).

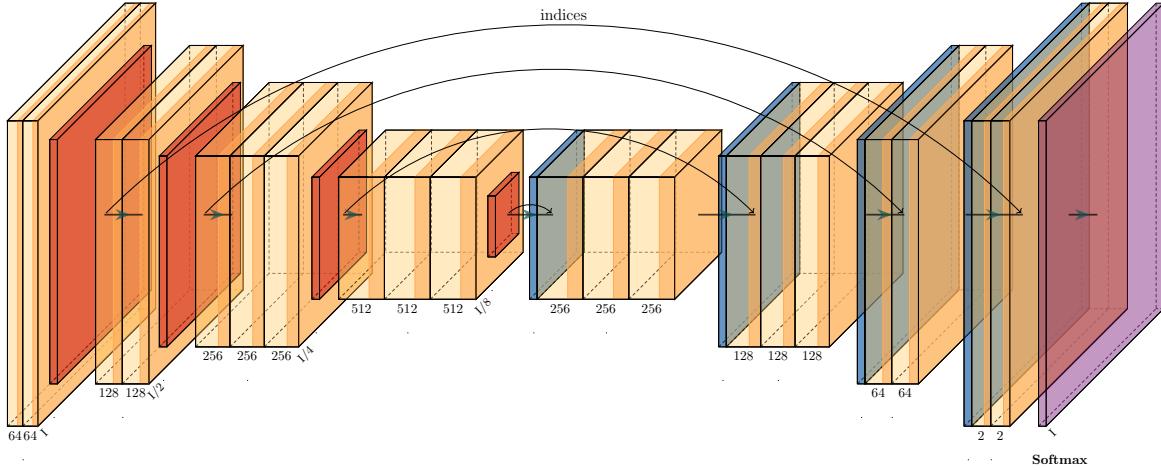


Figure 7: SegNet architecture. ([Badrinarayanan et al., 2015](#); [Iqbal, 2018](#)).

2.4.3 SegNet

SegNet iterates on the U-Net architecture by removing the transfer of full feature maps from encoder to decoder section, which uses large amounts of memory and contributes to the computational expense of running the model. Instead, the maxpooling layers pass the indices from which the pooled values are taken to the upsampling layer. This suggests a symmetric process of encoding and decoding, so the spatial understanding of ‘where’ the encoding is can be passed. SegNet, like FCN uses the VGG pre-trained convolution layers, but forgoes the skips, and instead adds convolution in a U-Net fashion to the decoder section. ([Badrinarayanan et al., 2015](#)). The variant SegNet-Basic was implemented for this work, as it is the most memory efficient but shares the same architecture and provides reasonable performance, only reducing the number of encoders and decoders by one. As our image sizes are small, this will likely not affect accuracy.

2.4.4 Activation Function

To convert the model outputs from the arbitrarily scaled logits produced by the model to a binary mask, we need to convert the logits to a probability. [Fazekas et al. \(2017\)](#) developed their model architecture to output a single logit value (One vs All), which can be more efficient when the number of samples of one class is limited ([Rifkin & Klautau, 2004](#)). By converting the task to a ‘multi-class’ architecture, the model now outputs logits which are related to the confidence that a given pixel is a class. We can therefore forego introducing thresholding at this step, and instead simply use the maximum of the two class outputs. This is a commonly used technique for binary segmentation ([Minaee et al., 2020](#)), and

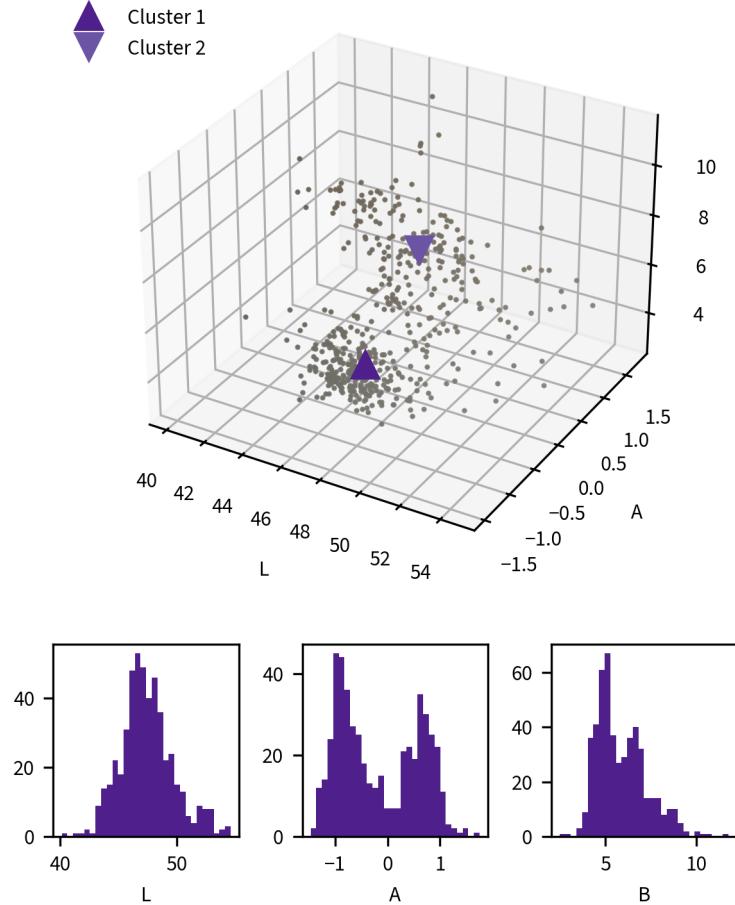


Figure 8: Plot of the mean colour values for all core sections in our dataset over CIELAB space. Means of the two main clusters are shown. Histograms show distribution of mean colour values.

removes an assumption of bimodal classes.

To convert logits to probability, we used the softmax activation function, defined in Eq. 6.

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (6)$$

where z is the logits output from the model, and K is the number of classes.

2.5 Data Augmentation

In order to increase the size of our labelled dataset, a number of augmentation transformations were applied to labelled patches when used in training. Two types of colour transforms were trialed over the data in the CIELAB space.

The Gaussian (GN) augmentation applies a random colour shift across the whole patch, selected from a normal distribution with a standard deviation of a fraction of the standard deviation of the colour channel. For GN augmentation, the number after indicates the noise fraction. For example, (GN0.5) indicates the standard deviation of the noise is 0.5 of the noise in the full dataset. This simulates extra data which was collected with the same structure but is a slightly different colour; as colour varies downcore, this may allow the model to generalise more easily.

The Cluster Switching (CS) method relies on an interpretation of the colour data. As the colour data is roughly bimodal (Fig 8), CS computes the nearest cluster, then shifts the colour until the mean is in the same relative location to the other cluster mean. This makes the assumption that the two clusters are similar in distribution, which roughly holds for our dataset. Augmenting in this way aims

to allow disturbance encodings learnt on one cluster to be applied to the other. This method is novel, and generates data with the most similar colour to the rest of the set. The two clusters are mostly distinguishable by redness.

Random vertical flip (RF) was also applied in some model runs. Horizontal flip was not used as the direction of gravity/drilling is important for the shape of some features. Rotation was not applied for similar reasons of directionality. Models which have had augmentation applied are indicating by appending the name of the augmentation strategy. When the augmentations were applied, each time an image was selected, there was a 50% chance of applying the augmentation. This means that the model was trained on a different set of augmented images each time, in the aim of increasing model generality.

2.6 Training

Loss was measured during training using the cross entropy loss due to its trivial differentiability, as is common in segmentation tasks (Badrinarayanan et al., 2015; Ronneberger et al., 2015). The cross entropy loss is measured as in Eq. 7.

$$\text{CEL} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log(\hat{y}_{ij}) \quad (7)$$

where N is the number of samples, C is the number of classes, y_{ij} is the true label for sample i and class j , and \hat{y}_{ij} is the predicted probability for sample i and class j .

Since validation loss does not require differentiability, we can use the more easily interpreted Intersection-Over-Union (IOU) metric, also known as the Jaccard Metric, defined simply as the ratio between the intersection of the predictions and ground truth labels and their union, without double counting the intersection (Eq. 8) (Jaccard, 1901).

$$\text{IOU} = \frac{y \cap \hat{y}}{y \cup \hat{y}} \quad (8)$$

where y is the true labelled disturbance, and \hat{y} is predicted disturbance by the model

The corresponding IOU Loss is defined as the negative of the IOU, and can be plotted alongside the CEL to evaluate model convergence and overfitting (Sec. 3.1). In cases where background removal is applied, we define two metrics, the Patch-IOU, which is the mean of the IOU over all of the patches in the test set, and the Image-IOU, which also averages over all of the removed patches by background removal, treating all of them as having an IOU of the background removal accuracy. This provides a comparable metric to the model IOUs from models where no background removal was applied.

Accuracy was also measured, defined as the ratio of correctly predicted pixels to the total number of pixels.

Training was conducted using a learning rate of 0.001, using the commonly used optimiser Adam (Kingma & Ba, 2015). A test train split of 0.6/0.4 was chosen. In order to better update the model, the full training set was iterated through several times. Each full iteration through the training set is known as an 'epoch' (Blarr et al., 2024; Long et al., 2015; Ronneberger et al., 2015). Most model runs were conducted to 3 epochs, as in this time, the models quickly decreased in loss, enabling us to see indicative results of the modifications applied. Since data augmentation mostly combats overfitting, the model runs with various augmentations were run for 20 epochs so we could obtain a plot of validation and training loss over time.

2.7 Post-Processing

The reference composite of Hodell et al. (2015) was used to stitch the full depth colour record, calibrated based on Ca/Ti ratio from XRF data and manually spliced. This makes the assumption that there was no stretch or squeeze, and that the surface colour data ties at the same point as the XRF; not always true in the presence of bioturbated surface features.

When averaging over the predicted cores for the generation of colour data, averages were only taken of rows in which at least 90% of the pixels remained. This number could be tested more robustly, but it was chosen as a good qualitative threshold, as if only a few pixels were not considered disturbance, they would have an outsize impact on the colour average, producing extra peaks/noise in the record.

2.8 Implementation

The models, training code and pre-processing, were implemented using the python pyTorch ([Paszke et al., 2019](#)) library for CNNs, which allows more control over model architecture than other libraries, and is faster to execute than Tensorflow/Keras for small models ([Novac et al., 2022](#)). Training was conducted on a single CUDA enabled NVIDIA GeForce 1650Ti with 4GB of vRAM. For the purposes of this report, all models were rewritten based on their architecture description so that they were implemented together using the same framework for more accurate comparison.



Figure 9: Segmentation on a representative core section (339_U1385A_01H_01). Epochs were chosen by peak performance, as tabulated in Tables 2 and 4.

3 Results

The models generated from our small labelled dataset was applied to the full unlabelled set of cores, and data generated from applying the model to the full composite section of [Hodell et al. \(2015\)](#) is given in Sec. 3.5. Model evaluation was done with the labelled dataset. In later subsections, representative core sections are shown, but training and validation was always done on the full set of labelled core sections.

3.1 Model Architecture

A number of training runs with various architecture and parameters were conducted to produce the best model (Table 2). To determine the optimal model architecture, we trained each model for three epochs to see the initial response and the compute time required for training.

All models decreased in loss over time, indicating that the optimiser and learning rate were appropriate. SegNet was both the fastest model to train, and produced the best IOU performance. UNet was slightly more accurate; this suggests that UNet produces more false positives. The lower batch sizes due to memory requirements seemed a major influence on the model training time. Qualitatively, it seems FCN8 did not have time to adapt to our training set (Fig 9). It segments some of the more obvious edges, which is likely thanks to the VGG backbone pre-training, but fails to encode anything further than that. In the core section chosen for Fig 9, it seems that both UNet and SegNet are able to segment most large cracks and voids, but struggle with the bioturbation features. SegNet is more convincingly

Architecture	Accuracy (%)	Patch-IOU	Training Time (min:sec)	Batch Size
FCN8	83.59	0.0099 ± 0.0041	42:03	5
UNet	93.06	0.1697 ± 0.032	29:08	8
SegNet	92.15	0.2203 ± 0.021	15:34	16

Table 2: Comparison of model architectures. All models were trained with the same parameters of 5x5 kernel for 3 epochs. Batch size was adjusted to fit the model within vRAM. Errors presented are the standard deviation of the IOU on the test set.

able to remove the soupy section near the coretop. In this section, SegNet segments more of the core than UNet, but UNet is more easily able to segment the narrow scrapes, likely due to being able to pass high resolution encodings.

3.2 Background Removal

SegNet, as the most efficient, was used to evaluate whether background removal improved the model IOU. To see the biggest difference, 128x128 patches with a 3x3 kernel were used. This was chosen to maximise performance on background regions, as the model should not need to recognise large features. The model was trained for 3 epochs. Without background removal, the model IOU was 0.420 ± 0.007 ; using only the cropped dataset, the model outputs an IOU of 0.195 ± 0.013 ; expected, as we've removed a lot of the 'easy' to predict segments. Using the corrected Image-IOU, this becomes 0.456 ± 0.013 . Errors reported are the standard error over 3 trials. These values are significantly different; background removal does improve model performance.

Background Removal	Patch-IOU Scores			Mean Patch-IOU	Mean Image-IOU
	Trial 1	Trial 2	Trial 3		
None	0.409	0.418	0.432	0.420 ± 0.007	0.420 ± 0.007
With	0.212	0.205	0.169	0.195 ± 0.013	0.456 ± 0.013

Table 3: Comparison of model performance with and without background removal. The model was trained for 3 epochs on 128x128 patches with a 3x3 kernel. Errors reported are the standard error across the three trials.

3.3 Data Augmentation

Four models, one without augmentation, and one with each of the three augmentation models were run for 20 epochs to see their effectiveness.

Augmentations	Peak Epoch	CEL	Accuracy(%)	Patch-IOU	Image-IOU	Training Time (h:min:sec)
None	9	0.269	92.3	0.226 ± 0.076	0.526 ± 0.076	5:34:12
GN0.1	2	0.435	91.8	0.229 ± 0.077	0.528 ± 0.077	5:27:09
CS	7	0.306	91.6	0.237 ± 0.089	0.533 ± 0.089	5:46:44
RF	3	0.379	83.3	0.232 ± 0.066	0.529 ± 0.066	5:34:50

Table 4: Epoch selection was done according to Fig. 10. Full image IOU was calculated as in Sec. 2.6. Augmentation strategies are in Sec. 4.2.3. CEL is the training loss. Errors are standard deviation of the IOU across the test-set at the given epoch.

Differences between models trained on augmented data are not significant, though the shape of the training curve indicates different degrees of overfitting (Fig. 10, Sec. 4.3.2). Overall, the Cluster Switching model performed slightly better at its peak.

The training and validation loss (CEL and IOU Loss) are shown per epoch in Fig. 10. The model with no augmentation, as expected, fits well for 10 epochs, and then begins to overfit; the validation loss increases while training loss continues to decrease. Cluster switching (CS) augmentation seems to slow overfitting slightly but introduces more error into the IOU calculation. Random flips increases the validation noise slightly, as does the gaussian noise augmentation. Overfitting seems to be reached similarly fast with RF, whereas GN slows overfitting significantly.

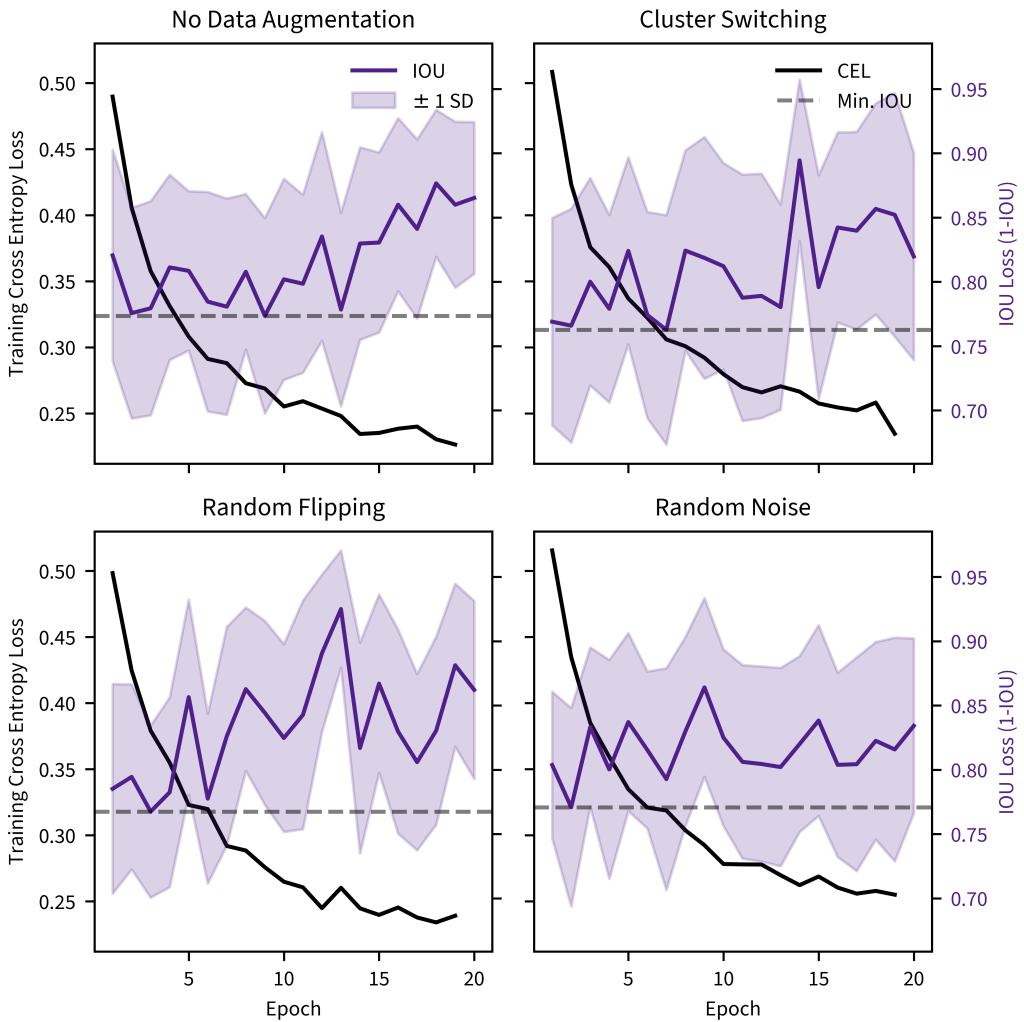


Figure 10: Training Loss and Test IOU Loss for different augmentation methods. Standard deviation is of the patch-wise IOU over the test set for one epoch of one model run.

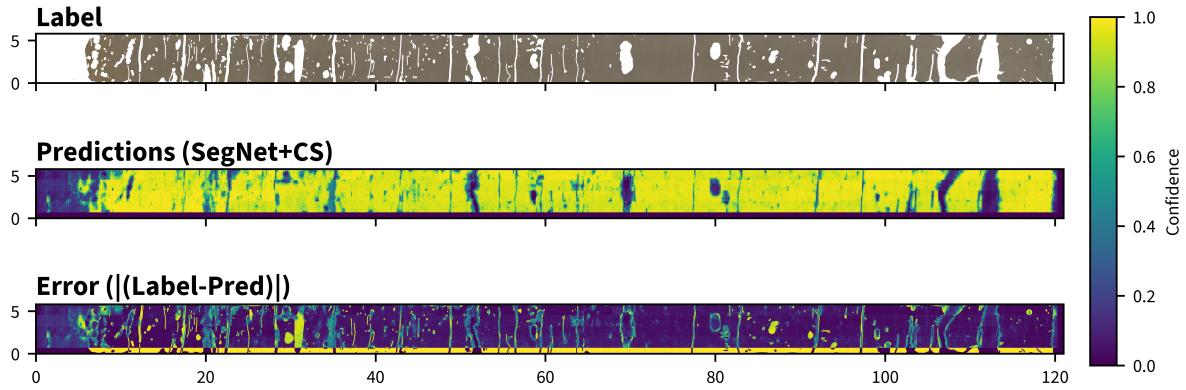


Figure 11: Visualisation of confidence down a core section. Confidence is defined as the probability of a pixel being undisturbed.

3.4 Confidence Maps

By applying the softmax activation function, we can obtain a map of how confident the model is that a given pixel is undisturbed (Fig 11). It is obvious that the model is more confident in large cracks and voids, and less confident in smaller features. Different confidence levels could be chosen for filtering the record (Sec 2.4.4). We also show the absolute error, (i.e. the absolute value of the label - confidence). The majority of errors are at the edge of disturbance features and outside of the patched area.

3.5 Filtered Colour Record

The best performing model from above, SegNet+CS was used to generate a full colour record versus depth on the composite section of [Hodell et al. \(2015\)](#).

In the full composite section (Fig. 12), the broad shape trends are the same as in the spectrophotometer record, and a number of sharp points are visible in both records. Errors in the unfiltered section are greatly reduced by the filtering, and all extreme artifacts are removed. The mean is roughly similar across the whole timescale. Detailed analysis of this full record was not carried out due to the constraints of this report.

3.6 Qualitative Comparison

Two core sections, U1385-A-01H-01 and U1385-A-13H-04 were chosen to view the model performance. A-01H-01 contains a number of cracks and voids, whereas A-13H-04 shows significant evidence of bioturbation.

Qualitatively, the model performs better on cracked sections, and sections of the redder sediment type. It is generally able to remove most voids and lithological accessories, as well as oxidation artifacts and some scrapes and scratches. Segmented images do have significant gaps in them; 23.8% and 25.8% of the image is segmented away. In Figure 13, large obvious cracks and voids are trivially removed, but there are also less obvious changes in the lightness record. Specifically highlighted in 01H-01 are dark spots from oxidation, deep gouges and scrapes, reflections from scrapes. There is a clear Thallasonoides burrow at 30cm downcore which is not segmented, and further burrows at 91cm.

In 13H-4, there is significantly more bioturbation, and it is notably worse segmented. Some burrows, such as the Planolites at 50cm and 144cm are selected, other burrows at 90cm, 120cm, 132cm and 138cm are well selected as well. Notably though, the light colour sediment at the top of the section is not well segmented, especially the Planolites and others from 0-42cm. This is likely as the training data does not contain a large amount of burrows in this lighter sediment.

There's more randomness and noise in the segmentation mask on 13H-4, which lines up with the higher probability of disturbance in cores that look heavily bioturbated, but the lack of specific masking across the whole core indicates this model is not perfectly tuned to select bioturbation. Highlighted panels in this figure indicate regions with well segmented larger burrows.

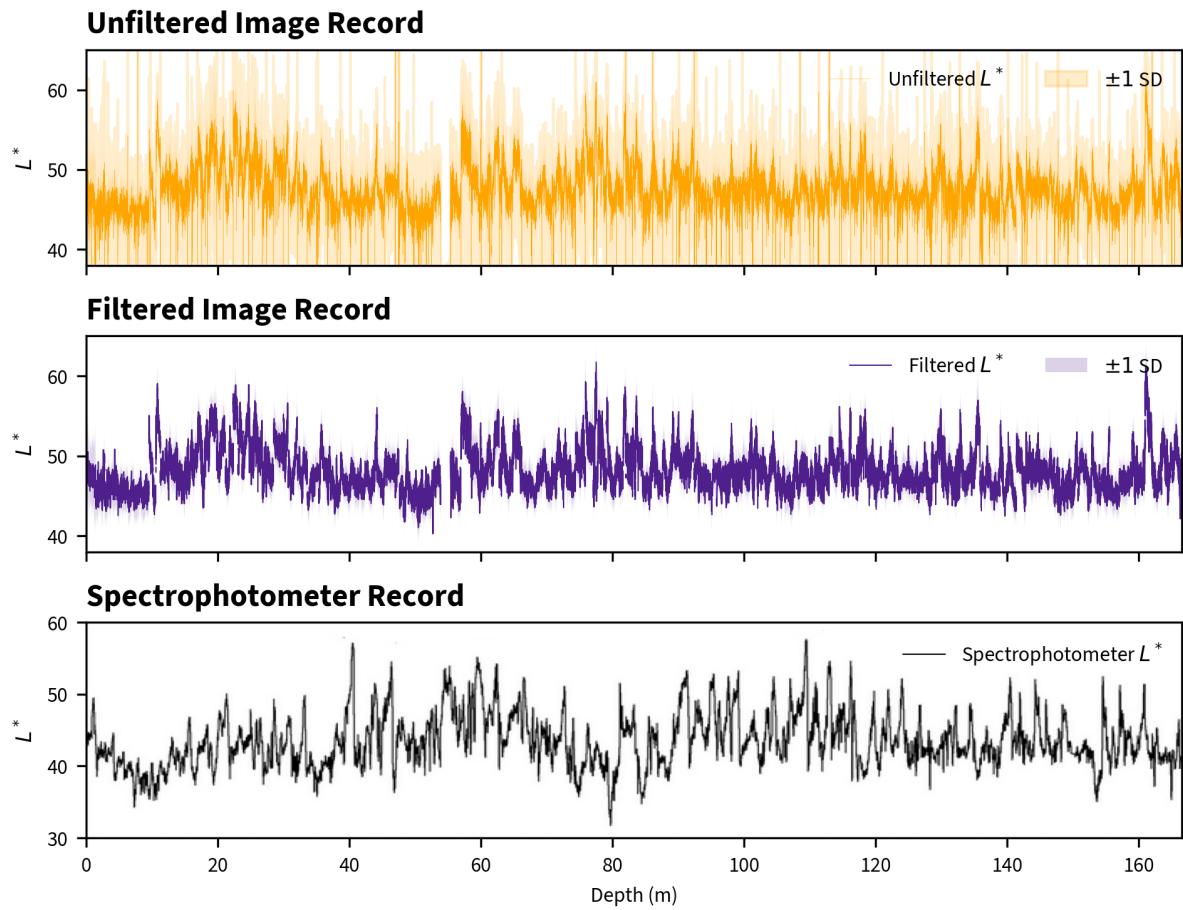


Figure 12: Unfiltered and filtered digital image record in L^* , plotted against a reference depth section. Gap at 55m is due to missing core in the image dataset available. Spectrophotometer record is taken directly from [Hodell et al. \(2015\)](#). Standard deviation in the deviation of the values averaged cross-core at each point.

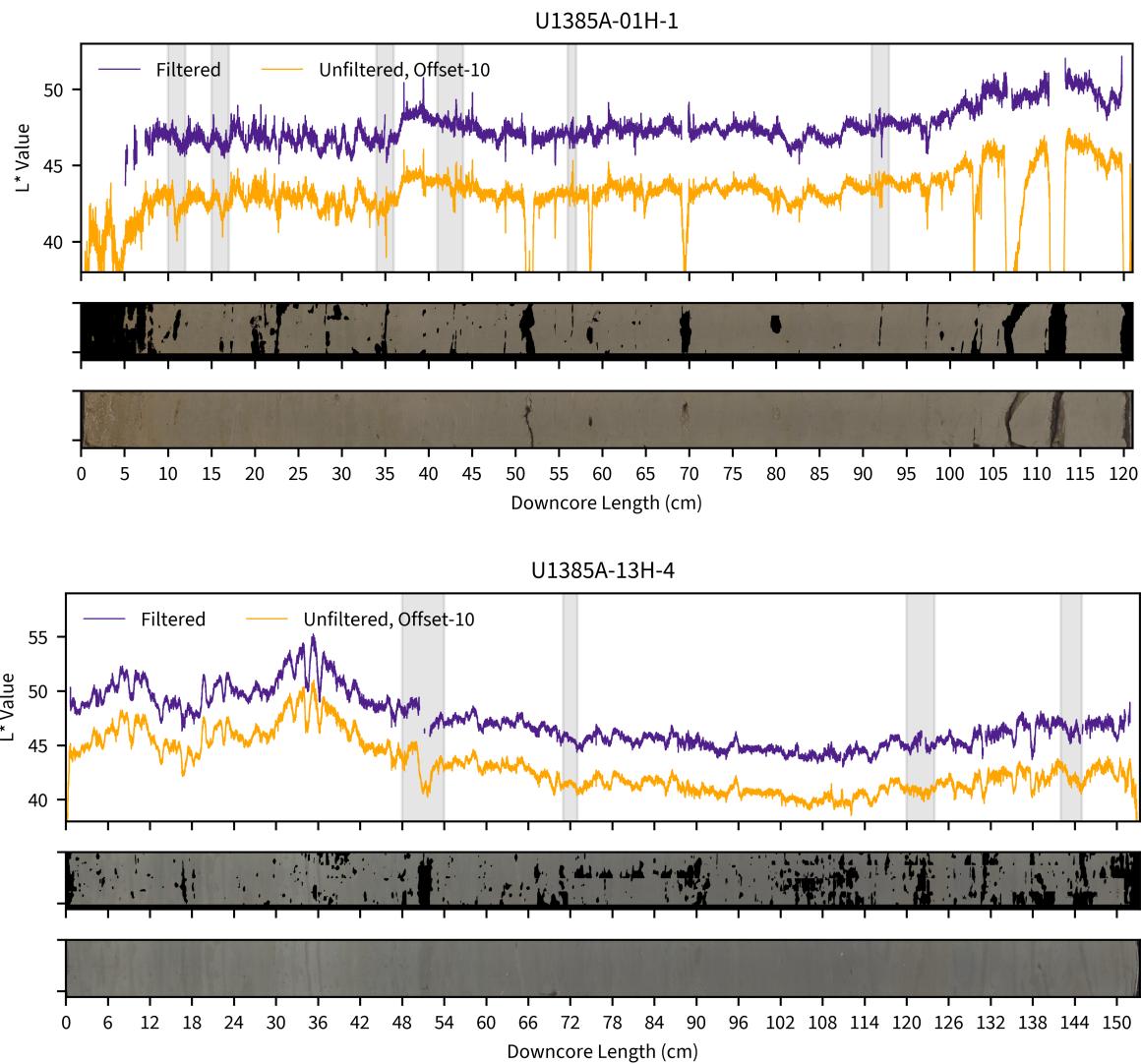


Figure 13: L^* lightness records from two core sections. Grey highlight shows regions where segmentation has changed the shape thanks to smaller disturbances.

4 Discussion

4.1 Method Usage & Generality

In this work we present a significantly improved method for automatically processing full cores. The developed `coreclean` package for python contains a full suite of scripts useable for this pipeline. All steps in this process are core, length and resolution agnostic, so segmentation and colour record generation can now be conducted much more quickly. Additionally, this method is implemented in pyTorch, is trivially deployable on cloud services and is extensible; the package serves as a starting point for further experimentation with CNN and algorithmic processing of core images in the quest for higher resolution of climate records.

4.2 Model Performance

Compared to the work of [Fazekas et al. \(2017\)](#), we produce a significantly higher IOU score of 0.53 ± 0.08 , and are able to train and segment full core sections much more quickly than the architecture used. The IOU score we receive is likely higher thanks to the relative ease of segmentation when the background of the core image is so large, as well as due to some architectural improvements. Our dataset of 256x256 images is smaller (4364 compared to 9993), and our training time of less than 5 hours per model is less than the previously reported 10 hours. Either our model is better at generalising quickly from small datasets or the U1385 cores are more easily segmented. Both of these factors may be true; the SegNet architecture was certainly more appropriate for our task based on the results in Tab. 2, though the background of our image was larger; meaning that it is easy to gain a boost to performance by applying background removal.

The model qualitatively selects most of the drilling disturbance, lithological accessories, voids and dark spots, especially where colour is significantly different. Bioturbation is less well segmented, likely because of the limited labelling in the dataset, though in some cases burrows are well removed.

4.2.1 Background Removal

Background removal was found to significantly improve model performance, as the model was able to use more of its parameters in encoding subtler information rather than just the background. This is likely more effective on our dataset due to the large backgrounds, as opposed to the work of [Fazekas et al. \(2017\)](#), where only a small ruler was present in the raw images. Our background removal technique, though it can require tuning to the exact shape of the core, is completely algorithmic and generalisable, and can automate 97.6% of core labelling (Tab. 1).

It is telling that the most significant improvement in model performance was not from the CNN itself, but instead from a simple algorithmic pre-processing step. Steps should be taken to ensure that we are not overcomplicating problems in image processing, as algorithmic methods are often less computationally intensive, and more prone to detectable and explainable errors.

4.2.2 Architecture

SegNet was chosen as the best performing model based on two factors. The first, computation time, in which the UNet and FCN model was significantly slower to train, because of the larger number of parameters and passing full maxpool layers through the model. Further, the choice of lower batch size thanks to the large memory requirements of both other models meant that the training was even slower than it could have been otherwise. Even so, both other models had worse performance after 3 epochs; likely because the dataset was small, and those models contain many parameters due to their high encoder depth, that they were not able to reach convergence with the data available. The pre-trained backbone of FCN is known to make it less receptive to small datasets ([Yang, 2024](#)).

4.2.3 Data Augmentation

As is visible in table 4, there is no significant improvement with data augmentation.

The random shift (GN) augmentation likely did not improve model performance as the average colour data in each patch is similar to any other, and even the validation set does not contain colours very far outside of the original model. Adding too much of this augmentation (i.e. a noise with standard

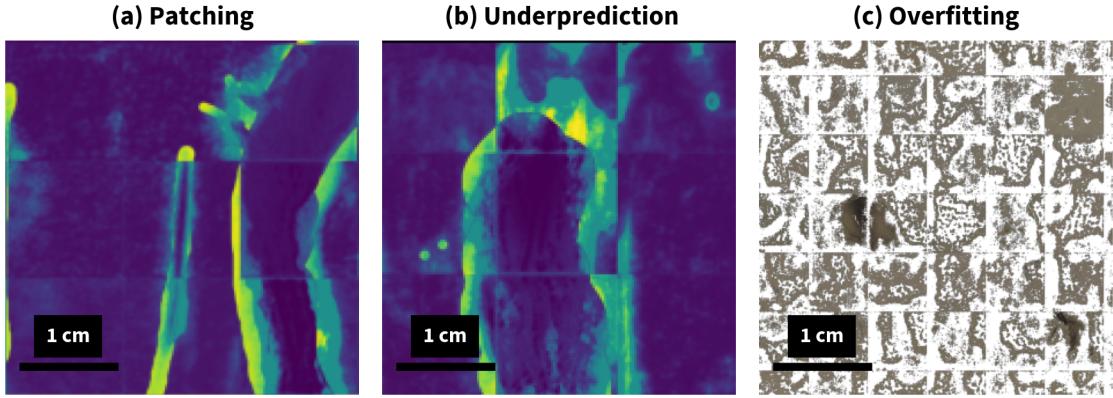


Figure 14: Three errors in the model. Panels a and b are taken from the error plot in Fig. 11. Panel c shows a masked image from a 128x128 SegNet model run for 20 epochs with kernel size 3.

deviation of 0.5 the between section standard deviation), causes a significant decrease in IOU, as we overwrite any information to do with colour variations in the model.

Random flip (RF) augmentation did not significantly improve model performance, though it is a commonly applied CNN augmentation. There are several possible reasons for this; one is that there is an asymmetry in the dataset, maybe ascribable to scraping direction or to lighting variations across the core. More likely, the patching process has an asymmetry due to the requirements for square patches of 128x128 or 256x256, which cuts off the bottom of the core. It's likely then, that the most disturbed sediment at the edge of the core is more present in the dataset at the top, and randomly flipping this does not make the model perform better in removing this. Suggested improvements to the patching process (Sec. 5.1.3), might mean that this augmentation becomes more useful.

Cluster switching (CS) augmentation was not found to significantly improve model performance either. One explanation possible for this is that the there is a difference in disturbance expression between colour clusters; i.e. the lighter sediment has more clear bioturbation (as organics generally leave a darker colour). This method did produce the model with the highest IOU score, so future work should investigate whether it does improve model fit when there is more labelled data available.

4.3 Model Limitations

4.3.1 Patching Artifacts

The process of patching the images into smaller sections for training and evaluation introduces some errors. Firstly, disturbance which spans multiple patches, especially when the sections in each patch are small (or do not have a distinctive shape), are generally not identified by the model (Fig. 14 (a)).

Once restitched, especially after more training epochs, often the edge of each patch is mispredicted. This is likely due to the implicit padding of 0 needed to be able to convolve a kernel over the patch. This specifically affects the power spectrum of the restitched image, providing an artifact signal with a wavelength of the patch length. This may be mitigated by more sophisticated padding types; specifically, a reflection padding or even padding by taking originally overlapping patches (Section 5.1.3).

4.3.2 Overfitting

In figure 10, after roughly 10-12 epochs, all but the GN model began to overfit. This is indicated by the validation loss increasing while the training loss continues to decrease. This indicates the model is recognising patterns unique to the training set, which is very possible with the large number of parameters in the model compared to our small dataset.

To demonstrate the structure of an overfit model, when a 128x128 patch model was left running for 20 epochs, overfitting selected seemingly random pixels which happened to minimise the loss on the training set, but had no relation to the generalisable structure of the image and disturbance. This is shown in figure 14 (c).

Our model has a significant problem with overfitting, beginning to reach this stage after only a few epochs, which meant that further training became useless. This is due to the small size of the labelled dataset, and its lack of variety. Most of the cores from the labelled set come from the top half of the section; and 9 of 12 demonstrate low levels of bioturbation. For improved performance on this, generation of a larger dataset so that training can contain more complexity for the model is recommended, but better tuning of data augmentation may also be able to supplement this.

4.4 Implications for Future Sampling

4.4.1 Taking Core Images

Historically, core sections have been photographed for reference and visual description only, rather than for the collection of high resolution colour data. As has been done before ([Nederbragt et al., 2006](#); [Wu et al., 2022](#)), and is done in this work, we now have the ability to produce a high resolution colour record limited only by the resolution and quality of the photographs taken. It is important for an undisturbed record that such photographs be taken as quickly as possible after cutting the core, as dehydration and exposure to the air provokes more disturbance in the form of cracking and oxidation of the sediment, but many of the segmented disturbances, specifically scrapes, are introduced by core preparation. To improve the resolution and amount of data in these digital image colour records, we suggest that it is possible to scrape the surface of the cores without leaving gouges ([Nederbragt et al., 2006](#)), as these produce sharp peaks in the unfiltered colour record (Figs. [12](#), [13](#)). In order to combat the impact of reflections which were present in many of our core images, it's important that lighting should be diffuse and that as much as surface water as possible is removed from the core surface; a drying process with warm air could be investigated to improve this quality.

4.5 Implications for Resolution of Colour Records

Qualitatively, all models seem to under-predict disturbance in core sections (Fig. [14](#), b), likely due to the small amount of labelled data. In the above figure, and in Fig. [11](#), most disturbances have a light rim, where the model has discarded less of the core than the human labeller. There are two interpretations for this; the first, is that the model has identified the feature and since it is labelling pixelwise, it can be more precise with the boundary; alternatively, the model is more confident when the kernel is within the sharp line in the center of the feature or fully covers the feature. Any distance greater than the kernel from an obvious disturbance is unable to be selected, and so it leaves slightly modified sediment away from the most obvious part of the feature.

Additionally, in Fig. [13](#), we can see that a number of bioturbational features are not segmented by the model, this is further underprediction. Despite this, our best SegNet-8 model already discards 26.6% of the core across the full composite, and leaves many gaps in the full record. This seems to suggest if colour data is to be used for robust climate predictions, more duplicate sections should be taken and combined, and more sophisticated modelling strategies should be used to combine several core sections.

Comparing this data to the justification for higher spectrophotometer resolution proposed by [Chapman & Shackleton \(1998\)](#), we apply differenced cross correlation to two core sections (Fig. [15](#), which can show us the information content lost by offsetting the record by small lags. The cross correlation was calculated between each half of the core, and differencing was applied to highlight changes.

Figure [15](#) shows that the cross correlation increases slightly for the segmented record. This suggests that the model is removing some smoothed noise, but this effect is small. When looking at the differenced cross-correlation, there is an increase in cross-correlation across all lags, which shows that we are removing distinctive features that are not similar to any other part of the core; the disturbance related features, especially the bioturbation features, (in 13H-4), were what caused the core sections to be less self-similar. The peak at 0 lag is visible in both sections, and it sharply drops off in 01H-1, suggesting a resolution of approx. 1cm is still present, but in the more bioturbated section (13-H-4), after segmentation, the peak is much less pronounced and wide. Most of the resolution here has been destroyed by the burrowing, and the high cross-correlation in the unfiltered section was only recording the position of the burrows themselves. At least in some sections, aiming for resolutions of greater than 1-2cm is impossible with a naive approach of segmentation.

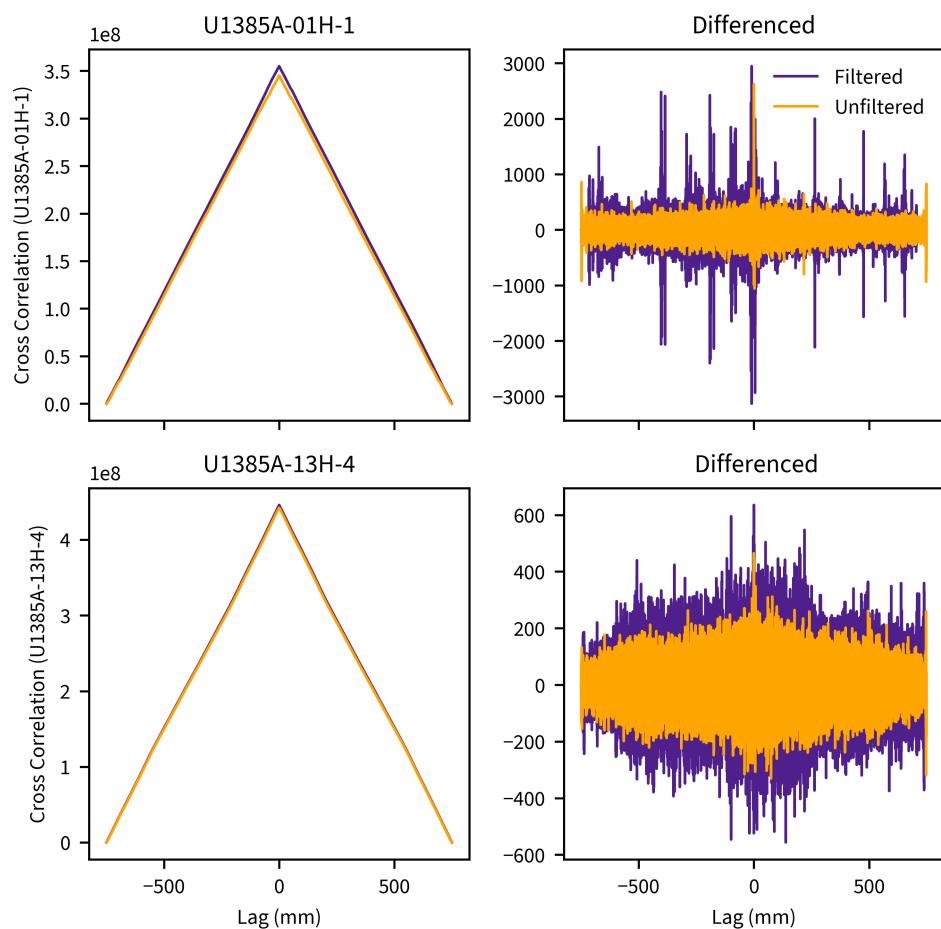


Figure 15: Cross correlation of two core sections, with and without differencing.

5 Conclusions

5.1 Future Work

5.1.1 Downsampling

One notable time optimisation which was not implemented within the scope of this report is the downsampling of patches for training and evaluation, and the application of the lower resolution mask generated to the full resolution core images. This would lower the training and usage demands of the CNN models, while sacrificing very little as the boundaries of most disturbances are not defined at pixel scale.

5.1.2 Alternative Augmentation

Data augmentation with small rotations could provide some improvement as most core disturbances can appear with small rotational offsets; technically, this is not possible while maintaining the maximum resolution during model training as it would require sub-pixel sampling; if downsampling was added, this would become feasible, as spatial transforms could be applied to the high resolution image and then downsampling would have more information to pull from.

5.1.3 Patching Improvements

Patching limits the maximum disturbance structure scale detectable by the model, specifically, the encoding of what the model sees as disturbance is only considered at a scale of a patch. This means that large scale disturbances, especially those that are similar in colour but have recognisable structures, are much harder to detect by our model. To combat this, a repeated structure such as presented by [Ronneberger et al. \(2015\)](#), where the patches are cropped too large by at least the kernel size and then fed to the model mean that the convolution structure could be applied seamlessly across the core.

Additional patching improvements could be made in the full core stitches; currently, non-full patches are discarded, which leaves small unpredicted lines across the bottom and at the end of the core. By cropping to integer multiples of the patch size only, or by selecting more of the image and then discarding extra patches in a post-processing step, we could improve model performance. Mirror padding could also help simply to avoid some patching errors.

5.1.4 Dataset Generation

As always with machine learning, the model can be improved by increasing the size and quality of the dataset. Further time could be invested in labelling these cores. There may be merit in following the work of [Dorador & Rodríguez-Tovar \(2018\)](#) and applying an additional colour thresholding procedure, and creating a rigorous dataset algorithmically on one set of cores. Data augmentation could then be applied to this dataset, to help generalise to more cores. Rather than manually labelling a dataset, if a simulated colour record could be generated from XRF or other data, training could be conducted against a true “Ground Truth” dataset, which would reduce the errors and time involved in labelling, as well as producing more robust metrics of model accuracy.

5.1.5 Ensemble Methods

With more time and compute available, the data augmentation strategies proposed and trialed could be rigorously compared for improvement, as without more model runs, the errors in IOU score can only be for a specific model training run. Almost all of the results of this report are simply indicative, and require a number more models to be trained for us to construct rigorous error bars on the efficacy of this technique.

5.1.6 Bayesian & Process Modelling

In ice core science, dating of proxy records is now commonly performed using the PaleoChrono package ([Parrenin et al., 2024](#)), which uses a probabilistic model to correct tie points and establish chronologies. It also allows us to correlate records to remove gaps; which is ideal for sediment records, as the disturbed areas create gaps, especially in high resolution records. Similar probabilistic tools should be used to

correlate the chronology and colour record from U1385, and could be combined with segmentation tools to produce a complete, high resolution record of colour.

Process modelling could help correlate different records, and address the more complicated question of where bioturbated material is sourced from, rather than simply segmenting it away.

5.1.7 Architecture Improvements

There are two different types of limitation of CNN based image processing. The first is the need for larger scale understanding of features. 5x5 kernels have been an improvement on 3x3 kernels [Fazekas et al. \(2017\)](#). This suggests that there are features with more spatial complexity than our model can recognise. Some of this can be solved by improved model structure, for example state of the art image segmentation often contains attention mechanisms and uses a transformer structure; obviously this requires large datasets and significant training time, but could produce better performance. A pre-trained large segmentation transformer model such as Meta's Segment Anything Model (SAM) ([Kirillov et al., 2023](#)) could be a solution to decrease training time, but it likely will not have strong accuracy to this problem without significant extra training, and cannot be hosted or run locally.

The problem of XAI (Explainable AI) is more structural ([Neha et al., 2024](#)). In sedimentology, the scientific validity of this method may be questioned as there is no explanation. Models of this type may begin to overpredict disturbance, as was the case for [Fazekas et al. \(2017\)](#). In this case, it becomes possible to generate a record that seemingly reduces the noise of colour signal, but discards valuable variation in the colour which contains information about past climate. Algorithmic checks and human verification is probably needed before machine-learning methods are applied to the generation of authoritative climate records.

5.2 Final Summary

In this report we have shown that colour reflectance data can be generated from digital images of cores, and that the record can be improved by the application of CNN based segmentation techniques. For low-compute applications and small datasets, we found that SegNet provides a more accurate and more well segmented result than FCN8 or UNet architectures, though this result is only indicative and may not be statistically significant if a large ensemble analysis is run. Background removal by morphological methods was found to significantly increase the IOU of our models based on 3 trials across the full labelled dataset of Iberian Margin core images. Three methods for data augmentation were trialed; two common, colour shifting and image flipping augmentations and one novel tactic which relies on the two main clusters of colours in the core sections within our dataset. None of these produced a significant improvement in model performance. Our most successful model, SegNet+CS, implemented the cluster switching data-augmentation strategy and obtained an IOU score of 0.53 ± 0.08 , which is an improvement on previous work by [Fazekas et al. \(2017\)](#), though this may be due to increased simplicity of background features in the dataset available for this work.

Qualitatively, this model was able to segment all drilling disturbance features and many other types of disturbance, but was not as consistent with bioturbation. Despite selecting some burrows accurately, a larger dataset is recommended for future training of models to improve the performance on more complex colour-agnostic features.

Our model segments 26.6% of the composite section, and likely underpredicts the disturbance in core sections. Despite this, cross-correlation analysis suggests that much of the resolution is no longer present in the section, which indicates that resolution higher than 1-2cm is likely destroyed by bioturbation, reflections and scrapes in some parts of the records. It is likely possible with an improved labelled dataset to segment more disturbance more accurately, and produce complete records of moderate resolution from parallel cores.

References

- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2015). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39, 2481–2495. <https://api.semanticscholar.org/CorpusID:60814714>
- Blarr, J., Kunze, P., Kresin, N., Liebig, W. V., Inal, K., & Weidenmann, K. A. (2024). Novel thresholding method and convolutional neural network for fiber volume content determination from 3d muct images. *NDT and E International*, 144, 103067. <https://doi.org/10.1016/j.ndteint.2024.103067>
- Bozzano, G., Kuhlmann, H., & Alonso, B. (2002). Storminess control over african dust input to the moroccan atlantic margin (nw africa) at the time of maxima boreal summer insolation: A record of the last 220 kyr. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 183(1), 155–168. [https://doi.org/10.1016/S0031-0182\(01\)00466-7](https://doi.org/10.1016/S0031-0182(01)00466-7)
- Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.
- Cai, H., Yang, Z., Cao, X., Xia, W., & Xu, X. (2014). A new iterative triclass thresholding technique in image segmentation. *IEEE Transactions on Image Processing*, 23(3), 1038–1046. <https://doi.org/10.1109/TIP.2014.2298981>
- Chapman, M. R., & Shackleton, N. J. (1998). What level of resolution is attainable in a deep-sea core? results of a spectrophotometer study. *Paleoceanography*, 13(4), 311–315. <https://doi.org/10.1029/98PA01067>
- Deaton, B. C., & Balsam, W. L. (1991). Visible spectroscopy: a rapid method for determining hematite and goethite concentration in geological materials. *Journal of Sedimentary Research*, 61(4), 628–632. <https://doi.org/10.1306/D4267794-2B26-11D7-8648000102C1865D>
- Dorador, J., & Rodríguez-Tovar, F. J. (2014). Quantitative estimation of bioturbation based on digital image analysis. *Marine Geology*, 349, 55–60. <https://doi.org/10.1016/j.margeo.2014.01.003>
- Dorador, J., & Rodríguez-Tovar, F. J. (2016). High resolution digital image treatment to color analysis on cores from iodp expedition 339: Approaching lithologic features and bioturbational influence. *Marine Geology*, 377, 127–135. <https://doi.org/10.1016/j.margeo.2016.02.005>
- Dorador, J., & Rodríguez-Tovar, F. J. (2018). High-resolution image treatment in ichnological core analysis: Initial steps, advances and prospects. *Earth-Science Reviews*, 177, 226–237. <https://doi.org/10.1016/j.earscirev.2017.11.020>
- Expedition 339 Scientists. (2013a). Methods. *Proceedings of the IODP*, 339. <https://doi.org/10.2204/iodp.proc.339.102.2013>
- Expedition 339 Scientists. (2013b). Site 1385 core descriptions. *Proceedings of the IODP*, 339. http://publications.iodp.org/proceedings/339/EXP_REPT/CORES/CORU1385.PDF
- Fazekas, S. Z., Obrochta, S., Sato, T., & Yamamura, A. (2017). Segmentation of coring images using fully convolutional neural networks. *2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 1–5. <https://doi.org/10.1109/ICITEED.2017.8250490>
- Gebregiorgis, D., Giosan, L., Hathorne, E. C., Anand, P., Nilsson-Kerr, K., Plass, A., Lückge, A., Clemens, S. C., & Frank, M. (2020). What can we learn from x-ray fluorescence core scanning data? a paleomonsoon case study [e2019GC008414 2019GC008414]. *Geochemistry, Geophysics, Geosystems*, 21(2), e2019GC008414. <https://doi.org/https://doi.org/10.1029/2019GC008414>
- Giosan, L., Flood, R. D., & Aller, R. C. (2002). Paleoceanographic significance of sediment color on western north atlantic drifts: I. origin of color. *Marine Geology*, 189(1), 25–41. [https://doi.org/10.1016/S0025-3227\(02\)00321-3](https://doi.org/10.1016/S0025-3227(02)00321-3)
- Haralick, R. M. (1992). *Computer and robot vision*. Reading, Mass. Addison-Wesley Pub. Co. <https://archive.org/details/computerrobotvis0000hara/page/n13/mode/2up>
- Hodell, D., Crowhurst, S., Skinner, L., Tzedakis, P. C., Margari, V., Channell, J. E., Kamenov, G., MacLachlan, S., & Rothwell, G. (2013). Response of Iberian margin sediments to orbital and suborbital forcing over the past 420 ka. *Paleoceanography*, 28(1), 185–199. <https://doi.org/10.1002/palo.20017>
- Hodell, D., Lourens, L., Crowhurst, S., Konijnendijk, T., Tjallingii, R., Jiménez-Espejo, F., Skinner, L., Tzedakis, P., Abrantes, F., Acton, G. D., Alvarez Zarikian, C. A., Bahr, A., Balestra, B., Barranco, E. L., Carrara, G., Ducassou, E., Flood, R. D., Flores, J.-A., Furota, S., ... Williams, T. (2015). A reference time scale for site U1385 (Shackleton site) on the SW Iberian Margin. *Global and Planetary Change*, 133, 49–64. <https://doi.org/10.1016/j.gloplacha.2015.07.002>
- Incarbona, A., Martrat, B., Di Stefano, E., Grimalt, J. O., Pelosi, N., Patti, B., & Tranchida, G. (2010). Primary productivity variability on the atlantic iberian margin over the last 70,000 years: Evidence from coccolithophores and fossil organic compounds. *Paleoceanography*, 25(2). <https://doi.org/10.1029/2008PA001709>

- Iqbal, H. (2018). Harisiqbala88/plotneuralnet v1.0.0. <https://doi.org/10.5281/ZENODO.2526395>
- Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des alpes et du jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37(142), 547. <https://doi.org/10.5169/seals-266450>
- Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In Y. Bengio & Y. LeCun (Eds.), *3rd international conference on learning representations, ICLR 2015, san diego, ca, usa, may 7-9, 2015, conference track proceedings*. <http://arxiv.org/abs/1412.6980>
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment anything. <https://arxiv.org/abs/2304.02643>
- Kittler, J., & Illingworth, J. (1986). Minimum error thresholding. *Pattern Recognition*, 19(1), 41–47. [https://doi.org/10.1016/0031-3203\(86\)90030-0](https://doi.org/10.1016/0031-3203(86)90030-0)
- LIMS reports. (2012). <https://web.iodp.tamu.edu/LORE/>
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- Marzeh, Z., Tahmasbi, M., & Mirehi, N. (2019). Algorithm for finding the largest inscribed rectangle in polygon. *Journal of Algorithms and Computation*, 51(1), 29–41. <http://jac.ut.ac.ir>
- Meijer, P. T., & Tuenter, E. (2007). The effect of precession-induced changes in the mediterranean freshwater budget on circulation at shallow and intermediate depth. *JOURNAL OF MARINE SYSTEMS*, 68(3-4), 349–365. <https://doi.org/10.1016/j.jmarsys.2007.01.006>
- Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2020). Image segmentation using deep learning: A survey. *CoRR, abs/2001.05566*. <https://arxiv.org/abs/2001.05566>
- Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, 16, 100258. <https://doi.org/10.1016/j.array.2022.100258>
- Nederbragt, A. J., Dunbar, R. B., Osborn, A. T., Palmer, A., Thurow, J. W., & Wagner, T. (2006). Sediment colour analysis from digital images and correlation with sediment composition [Cited by: 24]. *Geological Society Special Publication*, 267, 113–128. <https://doi.org/10.1144/GSL.SP.2006.267.01.08>
- Neha, F., Bhati, D., Shukla, D. K., Dalvi, S. M., Mantzou, N., & Shubbar, S. (2024). U-net in medical image segmentation: A review of its applications across modalities. <https://arxiv.org/abs/2412.02242>
- Novac, O.-C., Chirodea, M. C., Novac, C. M., Bizon, N., Oproescu, M., Stan, O. P., & Gordan, C. E. (2022). Analysis of the application efficiency of TensorFlow and PyTorch in convolutional neural network. *Sensors (Basel)*, 22(22).
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66. <https://doi.org/10.1109/TSMC.1979.4310076>
- Parkhi, O. M., Vedaldi, A., Zisserman, A., & Jawahar, C. V. (2012). Cats and dogs. *IEEE Conference on Computer Vision and Pattern Recognition*.
- Parrenin, F., Bouchet, M., Buzert, C., Capron, E., Corrick, E., Drysdale, R., Kawamura, K., Landais, A., Mulvaney, R., Oyabu, I., & Rasmussen, S. O. (2024). The paleochrono-1.1 probabilistic model to derive a common age model for several paleoclimatic sites using absolute and relative dating constraints. *Geoscientific Model Development*, 17(23), 8735–8750. <https://doi.org/10.5194/gmd-17-8735-2024>
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. <https://arxiv.org/abs/1912.01703>
- Rifkin, R., & Klautau, A. (2004). In defense of one-vs-all classification. *J. Mach. Learn. Res.*, 5, 101–141.
- Rodríguez-Tovar, F. J., & Dorador, J. (2014). Ichnological analysis of pleistocene sediments from the IODP site U1385 “Shackleton site” on the Iberian margin: Approaching paleoenvironmental conditions. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 409, 24–32. <https://doi.org/10.1016/j.palaeo.2014.04.027>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *ArXiv, abs/1505.04597*. <https://api.semanticscholar.org/CorpusID:3719281>
- Shackleton, N. J., Hall, M. A., & Vincent, E. (2000). Phase relationships between millennial-scale events 64,000–24,000 years ago. *Paleoceanography*, 15(6), 565–569.
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*.

- Thomson, J., Nixon, S., Summerhayes, C., Schönfeld, J., Zahn, R., & Grootes, P. (1999). Implications for sedimentation changes on the iberian margin over the last two glacial/interglacial transitions from ^{230}Th excess/0 systematics. *Earth and Planetary Science Letters*, 165(3), 255–270. [https://doi.org/10.1016/S0012-821X\(98\)00265-9](https://doi.org/10.1016/S0012-821X(98)00265-9)
- Wu, L., Huang, X., Cao, H.-s., & Wang, R.-j. (2022). Evaluating color parameters calculated from digital photographs of sediment cores as a tool in paleoenvironmental reconstruction—a case study using iodp site u1502 from the south china sea. *Sedimentary Geology*, 442, 106281. <https://doi.org/10.1016/j.sedgeo.2022.106281>
- Yang, H. (2024). Efficiency in constraint: A comparative analysis review of fcn and deeplab models on small-scale datasets. *Applied and Computational Engineering*, 75, 19–30. <https://doi.org/10.54254/2755-2721/75/20240501>
- Zhang, H., Huang, Y., Wijker, R., Cacho, I., Torner, J., Santos, M., Kost, O., Wei, B., & Stoll, H. (2023). Iberian margin surface ocean cooling led freshening during marine isotope stage 6 abrupt cooling events. *Nature Communications*, 14(1), 5390. <https://doi.org/10.1038/s41467-023-41142-8>
- Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, 57(4), 99. <https://doi.org/10.1007/s10462-024-10721-6>

Acknowledgements

I would like to thank the following people, without whom this work would not have been possible:

First, Dr. Stephen Obrochta and Dr. Szilard Fazekas for their advice in the direction and methods of this project, as well as for providing labelled core images from site U1085, and a pre-trained model from that site which was used in testing the patching and restitching code. No code was reused from their work.

My Director of Studies, Prof. Nicholas Butterfield for his academic support, and belief through the year.

Finally, thanks to my supervisor, Prof. David Hodell for his support, guidance, encouragement, patience, and knowledge throughout this project.

A Dependencies

The code was run in a conda environment with the following dependencies:

Package	Version	Package	Version
asttokens	3.0.0	blinker	1.9.0
Cartopy	0.24.1	certifi	2025.4.26
charset-normalizer	3.4.2	click	8.1.8
comm	0.2.2	contourpy	1.3.1
cycler	0.12.1	debugpy	1.8.11
decorator	5.1.1	executing	2.1.0
filelock	3.16.1	Flask	3.1.0
fonttools	4.55.8	fsspec	2024.12.0
graphviz	0.20.3	idna	3.10
imagecodecs	2025.3.30	imageio	2.37.0
ipykernel	6.29.5	ipython	8.31.0
itsdangerous	2.2.0	jedi	0.19.2
Jinja2	3.1.5	joblib	1.5.0
jupyter_client	8.6.3	jupyter_core	5.7.2
kiwisolver	1.4.8	largestinterrectangle	0.2.1
lazy_loader	0.4	lightning-utilities	0.12.0
llvmlite	0.44.0	MarkupSafe	3.0.2
matplotlib	3.10.0	matplotlib-inline	0.1.7
mpmath	1.3.0	natsort	8.4.0
nest-asyncio	1.6.0	networkx	3.4.2
numba	0.61.0	numpy	2.1.3
nvidia-cublas-cu12	12.4.5.8	nvidia-cuda-cupti-cu12	12.4.127
nvidia-cuda-nvrtc-cu12	12.4.127	nvidia-cuda-runtime-cu12	12.4.127
nvidia-cudnn-cu12	9.1.0.70	nvidia-cufft-cu12	11.2.1.3
nvidia-curand-cu12	10.3.5.147	nvidia-cusolver-cu12	11.6.1.9
nvidia-cusparse-cu12	12.3.1.170	nvidia-cusparselt-cu12	0.6.2
nvidia-nccl-cu12	2.21.5	nvidia-nvjitlink-cu12	12.4.127
nvidia-nvtx-cu12	12.4.127	opencv-python	4.11.0.86
packaging	24.2	pandas	2.2.3
parso	0.8.4	pexpect	4.9.0
pillow	11.1.0	pip	25.0.1
platformdirs	4.3.6	prompt_toolkit	3.0.48
psutil	6.1.1	ptyprocess	0.7.0
pure_eval	0.2.3	Pygments	2.19.1
pyparsing	3.2.1	pyproj	3.7.1
pyshp	2.3.1	PySimpleGUI	5.0.0.16
python-dateutil	2.9.0.post0	pytz	2025.1
pyzmq	26.2.0	requests	2.32.3
scikit-image	0.25.1	scikit-learn	1.6.1
scipy	1.15.1	seaborn	0.13.2
setuptools	75.8.0	shapely	2.1.0
six	1.17.0	stack-data	0.6.3
sympy	1.13.1	threadpoolctl	3.6.0
tifffile	2025.1.10	tiffile	2018.10.18
torch	2.6.0	torchmetrics	1.6.1
torchvision	0.21.0	torchviz	0.0.3
tornado	6.4.2	tqdm	4.67.1
traitlets	5.14.3	triton	3.2.0
typing_extensions	4.12.2	tzdata	2025.1
urllib3	2.4.0	wcwidth	0.2.13
Werkzeug	3.1.3		

B Code

Full code for the model pipeline is available in this [GitHub repository](#). The code is licensed under the MIT license.