Improving the feature network of EfficientDet on fossil detection

Wen Guan Gavin Chen, Dr. Benjamin Rosman, Dr. Steven James
School of Computer Science and Applied Mathematics
University of the Witwatersrand
Johannesburg, South Africa
2302891@students.wits.ac.za, Benjamin.Rosman1@wits.ac.za, Steven.James@wits.ac.za

***Abstract*—The excavation of fossils by palaeontologists has been insufficient due to the difficulties in determining the existence of the fossil from the sediment, Palaeontologists are looking for a model that automatically segments the bone fragments and fossils from muck images to improve the efficiency of exploring the fossils. Limited details of fossils are shown in the images. Using an object detection model might help detect the possible presence of fossils in an image, which can reduce time spent by avoiding the exploration of unnecessary areas and increase accuracy by pinpointing the exact location of the fossils. EfficientDet achieved the State-of-the-Art Performance in object detection. Based on EfficientDet, we modify the feature network to increase the capturing of fine details on the fossils in the image. We modified the feature network with a modified Feature Pyramid Network (FPN), Path Aggregation Network (PANet), Bi-directional Feature Pyramid Network (Bi-FPN), and Quad FPN. We also assess the fossil detection performance with the additions of the Repeating Block into the feature networks. We compared the feature network with the Repeating Block to the feature network without the Repeating Block on the fossil dataset using the metric of Average Precision with Intersection over Union (IoU). Our results demonstrate significant improvements in fossil detection accuracy with the addition of the Repeating Block in the modified FPN and PANet. However, for more complex feature networks like Bi-FPN and Quad-FPN, removing the Repeating Block leads to better performance. The modified Standard FPN with the Repeating Block and Quad-FPN without the Repeating Block produce a high AP of 34.36 and 34.33 on the fossil detection. Code is available at https://github.com/Gavin2417/Fossil-Detection.**

***Index Terms*—Fossil Detections, EfficientDet, FPN, Bi-FPN, Repeating Block**

I. Introduction

Palaeontologists have been able to reconstruct the history of life on Earth by examining fossils and bone fragments, unraveling the rich fabric of past ecosystems, and exposing the incredible diversity of species that once inhabited the world. Before the analysis, fossils and bone fragments must be discovered or unearthed from sedimentary rock formations, such as cliffs, quarries, and riverbeds. Caves and underground deposits often hold well-preserved fossils. Usually, larger-sized fossils and bone pieces can be unearthed faster than smaller-sized fossils and bone fragments, primarily because of their greater visible surface area. Smaller-sized fossils and bone pieces are mixed with the mud, stone, and other sediments. Palaeontologists have raised a concern that they manually separate those smaller bone fragments and fossils from the sediment as it requires a significant amount of background knowledge of the bone fragment for the separation [Sepkoski and Ruse 2015]. Particularly, certain smaller fossils and bone fragments may be mistaken for rocks or sand due to their similar shapes and colors. Due to the risk of damaging the fossils, this process is time-consuming and cautious, hence it requires a number of professional paleontologists to explore the fossils [Gong *et al.* 2007]. Therefore, having a model that can autonomously detect the fossils and bone fragments from muck images will assist palaeontologists in overcoming the issues. A muck image typically contains a mixture of mud, stone, bones, and fossils [Zhou *et al.* 2021]. Using an object detection model might help identify the possible presence of fossils in an image, which can save time by avoiding the exploration of unnecessary areas and increase accuracy by pinpointing the exact location of the fossils [Zhou *et al.* 2021].

There are several challenges faced in creating an object detection model on the muck images. Fossils and bone fragments are overlapped by mud or different-sized rock chips [Zhou *et al.* 2021]. Therefore, limited details of the fossils can be shown from the image, and even palaeontologists are not able to determine the fossils from those details of the image [Liu and Song 2020]. The sides of the rocks or fossils exhibit distinct structural surface planes [Zhou *et al.* 2021]. This caused classification errors and it classifies a fossil into two fossils. Due to the constant discovery of new fossils, there is a lack of data on images of new fossils and bone fragments for models to learn from [Kobayashi *et al.* 2021; Ruff *et al.* 2022].

There are not a lot of attempts in a similar field for fossil detection. A similar attempt has been made to identify fossils in rock images by applying UNet with ResNet for segmenting visible features in the rock [Liu and Song 2020]. MSD-UNet segmented the majority of the rock chips and provided characteristics of the rock chips such as length and area, for the geographer to evaluate the rock mass's conditions [Zhou *et al.* 2021]. Paleontologists only require the location of fossils or bone fragments within an image to excavate the fossils.

Convolutional Neural Networks (CNN) are commonly used for image segmentation tasks and object detection tasks since the network is able to learn spatial hierarchies of features in images [Long *et al.* 2015]. You Only Look Once (YOLO) used a single Neural Network for predicting the bounding box for the object and the object's class probability and achieved real-time detection while maintaining high accuracy and efficiency [Redmon *et al.* 2016]. EfficientDet has achieved state-of-the-art performance in object detection tasks on popular benchmarks such as COCO (Common Objects in Context) and Pascal VOC (Visual Object Classes) [Tan *et al.* 2020]. It outperforms most object detection models due to the shared features within the feature network [Zhang *et al.* 2022].

The feature of the object is essential for object detection. Therefore, the shared features' capability in the feature network from EfficientDet could be advantageous for fossil detection. The Repeating Block in the Bi-FPN plays an important role in EfficientDet that enables the feature network to process multiple iterations [Tan *et al.* 2020]. We aim to address the challenges posed by overlapping mud and rock fragments that often obscure the visibility of fossils.

In this report, we propose a modification to the feature network within the EfficientDet framework to enhance the capturing of fine details in fossil images. We explore the integration of various feature network architectures, including modified FPN, PANet, Bi-FPN, and Quad-FPN, to improve the model's performance. Additionally, we assess the impact of incorporating a Repeating Block in these networks to further improve fossil detection accuracy. We evaluate the modified models on the detection of fossils and bone fragments within sediment images, using the Average Precision (AP) metric.

This structure of the document is shown as follows: section II will introduce the concepts and the structures of EfficientDet and the different feature networks that are used for the modification. section III demonstrate the related works for this research report such as object detection models and some image segmentation model that are applied in the fossil field. section IV details the experimental setup with different feature networks used. It also shows the evaluation metrics used for model assessment. section V details the choice of the dataset, and the training process and provides a discussion of results and observations of the fossil detections. Lastly, section VI summarizes this research report and recommends further research and potential applications of the modified feature networks.

II. Background

EfficientDet consists of a backbone of EfficientNet, a Bi-directional Feature Prymaid Network (Bi-FPN), a class prediction network, and a box prediction networks. A representation for EfficientDet is shown in Fig. 1.

A. *EfficientNet*

The EfficientNet is tasked with extracting features at various spatial resolutions. This is crucial because objects in images are presented in different sizes. By extracting features at different spatial resolutions, the model can gather information about objects of varying scales, including the finer details of smaller objects. From the EfficientNet backbone of Fig. 1, each level of the Pyramid represents the resolution feature map. As the level increased, the resolution decreased by the downsampling of the previous level. The highest level (P_7) has the lowest resolution feature map compared to other levels but its feature map is able to capture higher semantics about the objects within the image [Meng *et al.* 2022].

B. *Feature Network*

Feature networks effectively leverage features from different levels of EfficientNet for the detection of the object by sharing or combining features across these levels. It reduces the risk of information loss because it does not rely on the features from each level from the EfficientNet only. So now, P_3 has the highest resolution feature map but also has the semantic information of other feature levels from the fusion, as is shown in Fig. 1. There are different feature network architectures used in various models to manage the fusion of features from different levels of EfficientNet.

1) FPN: Feature Pyramid Network (FPN) was originally designed as a top-down pathway in the lower feature level that will fuse the features' information from the higher feature level and the output from the convolution of the lower feature level by using element-wise addition [Gong *et al.* 2021]. This solved the issue of different scaled objects appearing in the images that the model can not capture the tiny objects' information from the image [Lin *et al.* 2017a].

2) PANet: Path Aggregation Network (PANet), integrates FPN with the bottom-up approach that also allows higher feature levels to contain features from the lower feature level instead of a top-down pathway only. The bottom-up approach is used to produce higher-resolution features at every feature level [Liu *et al.* 2018].

3) Bi-FPN: Bi-directional Feature Pyramid Network (Bi-FPN), is the development of NAS-FPN [Ghiasi *et al.* 2019] with repeated bi-directional approaches (top-down and bottom-up approach) to enable more high-level feature aggregation [Zhang *et al.* 2022]. It also removes nodes without feature fusion, those nodes do not have a high contribution to the feature network and it adds computational cost. Bi-FPN introduced a better weight feature fusion that combined both the bidirectional cross-scale connections and the fast normalized fusion techniques [Tan *et al.* 2020]. Bi-FPN used fewer parameters, and fewer floating-point operations per second (FLOPs) to achieve higher accuracy and efficiency [Tan *et al.* 2020].

4) Quad-FPN: Quad FPN extends the idea of Bi-FPN with PaNet [Raj *et al.* 2020]. The structure of Quad-FPN merges the feature level of a top-down approach with a bottom-up approach and the feature level of a bottom-up approach with a top-down approach. The structure is depicted in Fig. 2. This structure aims to improve the accuracy of object detection, particularly for objects with irregular shapes.

C. *Class Prediction Networks and Box Prediction Network*

The class prediction network handles the likelihood distribution across various object classes, whereas the box prediction network performs regression on the bounding box coordinates. The weights of the class and box network are shared across the levels of features.

D. Repeating Block

The concept of Repeating Block was first introduced by Neural Architecture Search Feature Pyramid Network (NAS-FPN), where it used an irregular feature network topology to reduce the computational cost while maintaining feature network performance [Ghiasi *et al.* 2019]. Then it was utilized in the Bi-FPN. The repeating block allows the feature network to iterate a specific number of times which helps in capturing enough information at various scales. The information at various scales is crucial for object detection.

III. Related Work

A neural network model was present that identifies the patterns in the model and recognizes simple, written characters with small variants [Fukushima 1980]. First CNN architecture was developed based on the neural network model. CNN is able to recognize more general handwritten characters than the neural network model because CNN is able to learn complex patterns [LeCun *et al.* 1998]. The development of computer vision started with the introduction of CNN on image classification [Krizhevsky *et al.* 2012]. Object detection involves the localization and identification of objects in an image. Object detection models are separated into two-stage and one-stage object detection models.

A. Two-Stage Object Detection Models

There have been numerous models developed based on CNN, including R-CNN families and those models are the representation of the two-stage object detection models. R-CNN by Girshick *et al.* [2014] is the first deep learning-based object detection model. Region-based CNN (R-CNN) first produced region proposals (RoI), then extracted features from feeding the RoIs, and lastly classified features into different categories [Girshick *et al.* 2014]. The use of sliding windows in the production of RoI slows down the system but it achieved high accuracy. Then, Fast R-CNN by Girshick [2015] was developed from R-CNN with RoIPool and a single network, which has a huge improvement in speed and accuracy. However, the process of generating Region of Interest (RoI) still significantly affects the overall processing time of Fast R-CNN. To overcome this issue, Ren *et al.* [2015] introduced Faster R-CNN with Region Proposal Network instead of the sliding window method for the regions in an image that potentially contain objects. Faster R-CNN achieved greater performance compared to R-CNN and Fast R-CNN [Ren *et al.* 2015].

B. One-Stage Object Detection Models

Redmon *et al.* [2016] proposed a one-stage detector system (YOLO) that determines the bounding boxes and class probability from a single network. Due to the system using the features of the whole image to make predictions; the background errors have decreased, and it improved its speed rapidly [Redmon *et al.* 2016]. Then, Tan *et al.* [2020] introduced a more accurate detection model (EfficientDet) than YOLO since YOLO was implemented for real-time detection. One of the main features of EfficientDet is the weighted Bi-directional Feature Pyramid Network (BiFPN) for combining the features at different resolutions, which is discussed in section II. While this model achieved high accuracy, it also used less computation and fewer parameters, which increased the training process and lowered the requirements of inputs [Tan *et al.* 2020].

C. Image segementaion

Prior to the introduction of image segmentation models, image segmentation used traditional methods - thresholding, regions-based, clustering, watershed [Basavaprasad and Ravindra 2014]. These techniques are determined from images' pixel intensities, distribution of intensities, edges, and gradient changes to separate the image into different regions. Those methods failed to achieve high accuracy, as Zhou *et al.* [2021] also showed that Multi-threshold and Watershed do not have correct segmentation on the rock. This is caused by low-contrasted images, the difference in the intensities value did not have a huge impact on the segmentation. One of the first semantic segmentation network proposed by Long *et al.* [2015], is a Fully Convolutional Network (FCN) that output a segmented image's map. FCN was inspired by CNN and changes all fully connected layers to fully convolution layers [Long *et al.* 2015]. Compared to the traditional method, FCN achieved better accuracy and speed due to the deep learning base network which learns rich and important features from the image.

There haven't been many attempts to detect fossils in this field, likely because there is limited published data available. An attempt has been made to identify fossils in rock images by applying UNet with ResNet for segmenting visible features in the rock [Liu and Song 2020]. UNet by Ronneberger *et al.* [2015] is based on FCN and adapted for biomedical segmentation purposes. This is a semantic segmentation model that contains an encoder (Down-sampling) and decoder (Up-sampling) with skip connections. The skip connections enable the model to maintain more features from the input picture into the output's segmentation map because down-sampling caused finer details lost during the process [Zhou *et al.* 2021]. MSD-UNet segmented the majority of the rock chips and provided characteristics of the rock chips such as length and area, for the geographer to evaluate the rock mass's conditions [Zhou *et al.* 2021]. The MSD-UNet model failed to segment the miniature rocks and the rocks that have less information presented in the image [Zhou *et al.* 2021].

IV. METHODOLOGY

We aim to compare the effectiveness of the repeating block on the FPNs in detecting the fossils and bone fragments from the muck image. We propose the modification of the feature network on the EfficientDet Model to enhance the feature

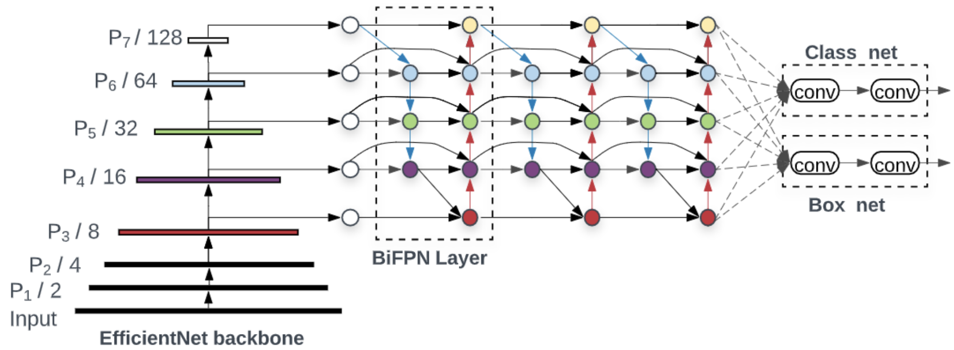

Fig. 1. **EfficientDet architecture** [Tan *et al.* 2020] – It used EfficientNet as the backbone network, BiFPN as the feature network, and a shared class/box prediction network. The Repeating Block is used in the feature network and the class/box prediction network.

capturing. Specifically, we adapt the feature network using various FPN structures, including the modified standard FPN, PANet, Bi-FPN, and Quad-FPN. We conduct a comparison between the four various FPN architectures, first without any repeating block, and then with the addition of a Repeating Block.

Originally, FPN was designed with a top-down pathway, where information from higher feature levels (lower spatial resolution) is transmitted to lower feature levels (higher spatial resolution) [Lin *et al.* 2017a]. In our modified Standard FPN, we extend the standard FPN architecture by incorporating an additional convolution after following the standard FPN approach. This convolution step aims to augment feature learning and enable the capture of finer details at each level of the image. The different feature networks without the Repeating Block are illustrated in Figure 2 and the different feature networks without the Repeating Block are illustrated in Figure 3.

We evaluate the models using the Average Precision (AP) with Intersection Over Union (IoU). This metric assesses how well the ground truth bounding boxes and the predicted bounding boxes are aligned. AP is a metric used to assess the accuracy of an object detector over a range of recall values. Intersection over Union (IoU) measures the intersection area between the predicted bounding box and the ground truth bounding box. It is calculated as the area of overlap divided by the area of union between the two boxes. The IoU threshold specifies the minimum percentage of IoU required for a detection to be considered correct. A detection is classified as a true positive (TP) if its IoU is greater than or equal to the specified threshold; otherwise, it is categorized as a false positive (FP).

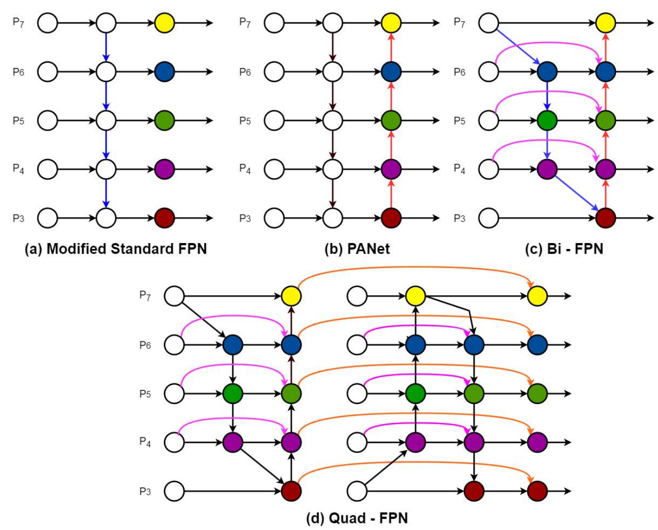

Fig. 2. **Feature networks design without the Repeating Block** - (a) is the modified standard FPN with top-down approach and an additional convolution; (b) is the PANet with top-down approach and bottom-up approach; (c) is the Bi-FPN with repeated bi-directional approaches (top-down and bottom-up approach); (d) is Quad FPN that merges the feature level of a top-down approach with a bottom-up approach and the feature level of a bottom-up approach with a top-down approach

V. Experiment and Result

A. Dataset

There are 131 images in the dataset that comprise of two types of images: images of isolated bone fragments against a clear background, as is shown in (a) of Fig. 4 and images depicting bone fragments mixed with sediments like sand and dust, as is shown in Fig. 5. Each image displays varying quantities of fossils along with distinct types of bone fragments. Some images contain more than 400 bone fragments. Every visible fossil or bone fragment will be annotated with a bounding box. All annotations are assigned the class label fossil. Augmentation techniques were implemented, including horizontal/vertical flips, partial image cropping, and slight rotations. Then all images are resized into 512×512 pixels. The image dataset is separated into a training dataset, a validation dataset, and a testing dataset at 70 %, 20% and 10 % respectively. The image annotations, which store both the image filename and bounding box details, is saved in the TensorFlow object detection CSV format.

We trained the model using images of isolated bone fragments against a clear background with the annotations, as shown in the image of Fig. 4. We used the trained model to

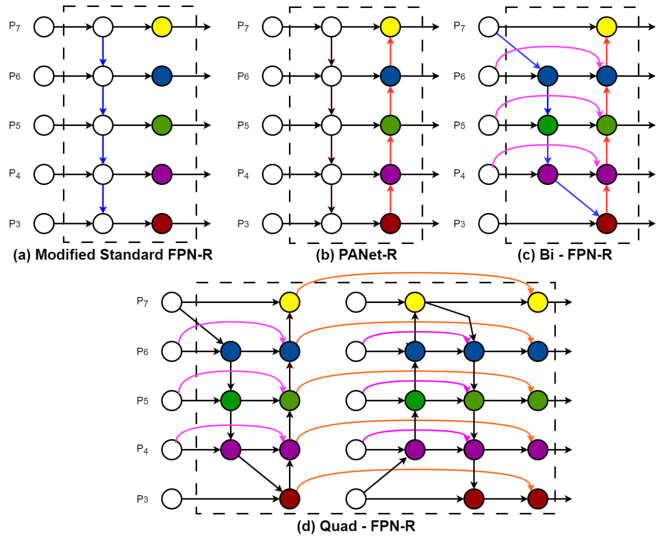

Fig. 3. **Feature networks design with the Repeating Block** - (a) is the modified standard FPN with top-down approach with the Repeating Block; (b) is the PANet with the Repeating Block; (c) is the Bi-FPN with the Repeating Block; (d) is Quad FPN with the Repeating Block

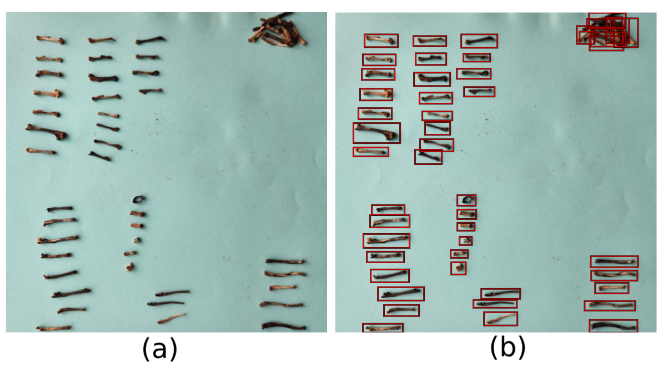

Fig. 4. **Initial training dataset** - (a) is the image of isolated bone fragments; (b) is the image that labeled the fossils with the bounding box

predict the presence of bone fragments in the images of bone fragments mixed with sediments. Due to the training from the clear background in the images of isolated bone fragments, the model struggles to accurately identify the majority of fossils within the sands and the stones from the images depicting bone fragments mixed with sediments, as is shown in the image of Fig. 5. The fossils and the bone fragments have similar colors and features to the background of the sediment. Therefore, the training, and validation dataset requires mixing both types of images: images of isolated bone fragments against a clear background, and images depicting bone fragments mixed with sediments like sand and dust, in order to provide a more accurate prediction of the fossils.

B. Training

The model takes in an image of size 512 $\times$ 512 pixels and outputs the image identification number, predicted boxes, and predicted class confidences as well as predicted class

Fig. 5. Fossil detection with the red bounding box using images of isolated bone fragments for training

labels. Although the EfficientDet-D7 model achieved 55.1% AP on the COCO test dataset, the EfficientDet-D3 model achieved 47% AP with much fewer parameters and fewer FLOPs (Floating-point Operations Per Second) [Tan *et al.* 2020]. So, the model is based on pre-trained TensorFlow EfficientDet3 with EfficientNet B3 with the format in Fig. 1. We modified the feature network of the EfficientDet with the modified standard FPN, PANet, Bi-FPN, and Quad-FPN.

Both Bi-FPN and Quad-FPN incorporate the repeating block. To conduct a fair comparison, we exclude this block from the Bi-FPN and the Quad-FPN. All feature networks, which are the modified standard FPN, PANet, Bi-FPN, and Quad-FPN, do not contain the Repeating Block. Then, we compare the result against the feature networks with the integration of the Repeating Block. We set the number of iterations for the Repeating Block to 3. We employ fast normalized fusion instead of additive fusion for all the models due to normal fast fusion offers better performance [Tan *et al.* 2020]. We apply the Focal Loss function to all models, which serves to emphasize the importance of the foreground region where objects are located [Lin *et al.* 2017b]. After the hyperparameter tuning process, the prediction confidence threshold is set to 0.3, the learning rate is set to 0.0024, and the Weighted Boxes Fusion IoU threshold is set to 0.62.

C. Results

We evaluate each model with the test dataset using metrics of $AP_{0.50}$, $AP_{0.75}$, and $AP_{0.50:0.95}$. The values 0.5, 0.75, and 0.50:0.95 represent the IoU thresholds. $AP_{0.50:0.95}$, which represented Average Precision over the range of IoU thresholds from 0.5 to 0.95 with a step size of 0.05, provides a thorough result.

According to the data presented in Table I, we can observe that both FPN and PANet when used with the Repeating Block, achieve higher overall AP accuracy compared to their counterparts without the Repeating Block. There is an approximate 2% increase for the accurate detection with IoU thresholds of 0.5 and a 1% increase for the accurate detection with IoU thresholds of 0.75. We can also observe that the absence of the

Repeating Block on the Bi-FPN and Quad-FPN increased the accurate detection and achieved better results on their original architecture. The Quad-FPN has a 5% Average Precision more than the Quad-FPN-R with the Repeating Block. From the TABLE I, it also indicates that FPN-R achieves similar accuracy to Quad-FPN. Bi-FPN achieves the highest on the $AP_{0.50}$. FPN achieves the highest on $AP_{0.50:0.95}$ as well as on $AP_{0.75}$.

The improvement in performance observed with the addition of the Repeating Block in the modified Standard FPN-R and PANet-R may be attributed to the inclusion of the repeating block enabling the model to enhance the features of the bone fragments and their surrounding environments. As shown in the results of TABLE I, one iteration of the modified Standard FPN and PANet proved to be insufficient, as it resulted in the model failing to capture enough details of the bone fragments. The three iterations of the modified FPN-R and PANet-R might lead to improved localization or detection of more fossils with higher Average Precision (AP) accuracy when compared to the modified FPN and PANet, as shown in TABLE I.

While the Bi-FPN shares a similar network architecture with PANet, which uses a top-down approach followed by a bottom-up approach, it distinguishes itself by eliminating nodes that do not significantly contribute to feature fusion. Additionally, it also fuses features from the initial representation in the EfficientNet with the output of each level at the end of the feature network. This is shown in the Fig, 2. Hence, the incorporation of the Repeating Block in Bi-FPN-R does not improve the detection of bone fragments as shown in Table I. This may indicate that the original network architecture of Bi-FPN, without the Repeating Block, captures sufficient fossil features across different scales in a single iteration. In contrast, the inclusion of the Repeating Block Bi-FPN-R seems to lead the model to overlearn these features.

Comparing the result of Quad-FPN and Quad-FPN-R from TABLE I, the higher accurate fossil detection results achieved by the Quad-FPN can be attributed to the absence of the Repeating Block. Consequently, the Quad-FPN without the Repeating Block effectively captures fossil features across varying resolutions of the image. Hence, incorporating the Repeating Block in Bi-FPN and Quad-FPN could potentially result in the model overlearning the features of fossils or bone fragments, leading to overfitting.

When employing the EfficientDet D3 model with various feature networks for fossil detection, the highest $AP_{0.50:0.95}$ achieved is 34.36%. In contrast, the EfficientDet D3 model achieves 47% $AP_{0.50:0.95}$ on the COCO dataset [Tan *et al.* 2020]. This notable difference may be explained by the fact that the fossil dataset contains examples of fossils or bone fragments that are difficult to identify even for humans, due to their similar colors and shapes to the background, whereas the COCO dataset comprises a wide range of large objects with significant color variations.

In Fig. 6, the detection of fossils using various feature network architectures is demonstrated. The top-left image (a) corresponds to the Modified FPN, the top-right (b) to PANet, the bottom-left (c) to Bi-FPN, and the bottom-right (d) to Quad-FPN. Each image is annotated with three different labels: the white-boxed label represents the ground truth annotations, the red-boxed label represents the annotations from the feature network without the Repeating Block, and the blue-boxed label represents the annotations from the feature network with the Repeating Block. The four images from Fig. 6, are the same with the same ground truth annotations.

TABLE I

COMPARISON OF FPNS WITH AND WITHOUT REPEATING BLOCK

Model	$AP_{0.50:0.95}$	$AP_{0.50}$	$AP_{0.75}$
Modified FPN	33.06	59.98	32.87
Modified FPN-R	**34.36**	**61.77**	**35.72**
PANet	32.65	60.04	30.81
PANet-R	**33.59**	**62.65**	**30.52**
Bi-FPN	**33.74**	**63.51**	**32.32**
Bi-FPN-R	32.69	62.32	29.77
Quad-FPN	**34.33**	**60.86**	**35.49**
Quad-FPN-R	32.32	61.48	30.24

Upon observation, it is evident that all models effectively identify the majority of the bone fragments. However, a notable challenge arises in the case of the bone fragment with the sole white bounding box in the middle of the images from Fig. 6, which is particularly challenging to discern due to its concealment within the sand.

Comparatively, the blue label detections provide a slightly better localization compared to the red label detections. However, it is worth noting that the inclusion of the Repeating Block also leads to the detection of areas that do not contain fossils. From the image (a) and (b) of Fig. 6, the red labels from FPN and PANet has more detection errors on the fossil compared to the FPN-R and PANet-R. Although the Bi-FPN-R has fewer detection errors, it also results in fewer detections on the area where the fossil exists, as shown in (c) of Fig. 6. Quad-FPN-R has made more detection errors compared to Quad-FPN. Towards the upper-middle edge of the image, the models detected a dark bone fragment, even though it was not annotated as ground truth. This highlights the models' ability to identify potentially important features that may have been overlooked in the initial labeling process.

VI. CONCULSION

This paper presents an approach for detecting the fossils and bone fragments among the sediments by using an EfficientDet model and adapting the feature network from the model with modified FPN, PANet, and Quad-FPN. We also evaluate the performance of integrating the Repeating Block on the different FPNs by using the metrics of Average Precision with Intersection Over Union. The results indicate a notable improvement with the incorporation of the Repeating Block into the modified FPN and PANet. This addition led to a significant improvement in both the modified FPN and PANet, enabling the model to capture intricate details of fossils across different scales and shapes. Removing the Repeating Block from both Bi-FPN and Quad-FPN leads to an improvement in fossil detection. This improvement can be attributed to

theoretical techniques for image segmentation. *Int. J. Comput. Appl*, 975:8887, 2014.

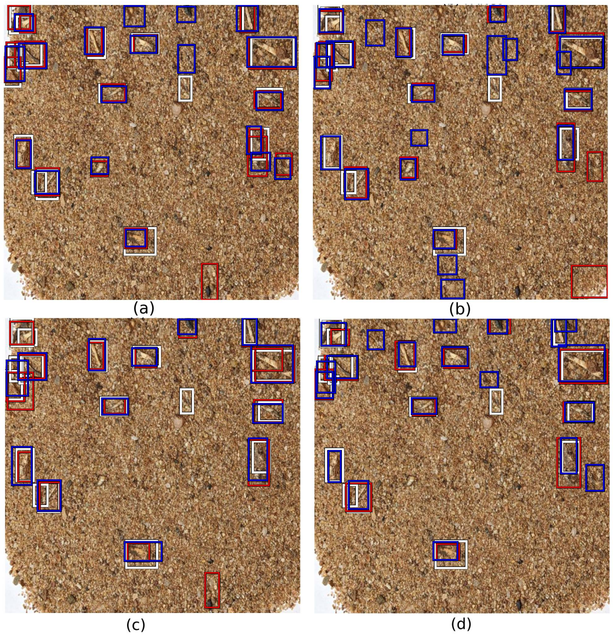

Fig. 6. **Comparison on fossil detections with different feature networks** - (a) used FPN architecture; (b) used PAnet architecture; (c) used Bi-FPN architecture; (d) used Quad-FPN architecture; The white bounding box is the ground truth annotation that contains fossils; The red bounding box represents the annotations from using the feature network without the Repeating Block; The blue bounding box represents the annotations from using the feature network with the Repeating Block

their inherent feature design, which is capable of capturing intricate details of fossils even without the need for the Repeating Block. We may conclude that the Repeating Block proves advantageous for simpler feature networks, but its impact on more complex feature networks is limited for the fossil dataset. The modified Standard FPN with the Repeating Block and Quad-FPN without the Repeating Block produce the highest accuracy on fossil detection. The modification of the feature network still cannot resolve the challenge of predicting faint or less distinct fossils that are not clearly visible. Furthermore, another possible improvement may arise for this fossil detections, it is included the attention on the EfficientNet and the feature network [Zhao and Wu 2019]. The attention helps the model dynamically focus on crucial regions of the input, aiding in the detection of fossils and bone fragments in the image.

Acknowledgment

My sincerest thanks to my supervisors Dr. Benjamin Rosman and Dr. Steven James for their guidance and support throughout this research. I would also like to thank Sandee Oster, for providing the dataset, sharing information about the fossils and bone fragments, and offering assistance in the process of labeling the dataset.

References

[Basavaprasad and Ravindra 2014] B Basavaprasad and S Hegadi Ravindra. A survey on traditional and graph

[Fukushima 1980] 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.

[Ghiasi *et al.* 2019] Golnaz Ghiasi, Tsung-Yi Lin, and Quoc V Le. Nas-fpn: Learning scalable feature pyramid architecture for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7036–7045, 2019.

[Girshick *et al.* 2014] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[Girshick 2015] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[Gong *et al.* 2007] Q.M. Gong, J. Zhao, and Y.S. Jiang. In situ TBM penetration tests and rock mass boreability analysis in hard rock tunnels. *Tunnelling and Underground Space Technology*, 22(3):303–316, 2007.

[Gong *et al.* 2021] Yuqi Gong, Xuehui Yu, Yao Ding, Xiaoke Peng, Jian Zhao, and Zhenjun Han. Effective fusion factor in fpn for tiny object detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1160–1168, January 2021.

[Kobayashi *et al.* 2021] Yoshitsugu Kobayashi, Ryuji Takasaki, Katsuhiro Kubota, and Anthony R Fiorillo. A new basal hadrosaurid (dinosauria: Ornithischia) from the latest cretaceous kita-ama formation in japan implies the origin of hadrosaurids. *Scientific reports*, 11(1):8547, 2021.

[Krizhevsky *et al.* 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[LeCun *et al.* 1998] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[Lin *et al.* 2017a] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.

[Lin *et al.* 2017b] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[Liu and Song 2020] Xiaokang Liu and Haijun Song. *Automatic identification of fossils and abiotic grains during carbonate microfacies analysis using deep convolutional* *neural networks*, 2020.

[Liu *et al.* 2018] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8759–8768, 2018.

[Long *et al.* 2015] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[Meng *et al.* 2022] Tianjian Meng, Golnaz Ghiasi, Reza Mahjourian, Quoc V Le, and Mingxing Tan. Revisiting multi-scale feature fusion for semantic segmentation. *arXiv preprint arXiv:2203.12683*, 2022.

[Raj *et al.* 2020] Shivam Raj, Natasha Sebastian, and Raj Bisht. Extradet with quad feature pyramid networks. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–6. IEEE, 2020.

[Redmon *et al.* 2016] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[Ren *et al.* 2015] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

[Ronneberger *et al.* 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[Ruff *et al.* 2022] Christopher B Ruff, Adam D Sylvester, Neni T Rahmawati, Rusyad A Suriyanto, Paul Storm, Maxime Aubert, Renaud Joannes-Boyau, Harold Berghuis, Eduard Pop, K Joost Batenburg, et al. Two late pleistocene human femora from trinil, indonesia: Implications for body size and behavior in southeast asia. *Journal of Human Evolution*, 172:103252, 2022.

[Sepkoski and Ruse 2015] David Sepkoski and Michael Ruse. *The paleobiological revolution: essays on the growth of modern paleontology*. University of Chicago Press, 2015.

[Tan *et al.* 2020] Mingxing Tan, Ruoming Pang, and Quoc V. Le. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[Zhang *et al.* 2022] Zhengning Zhang, Lin Zhang, Yue Wang, Pengming Feng, and Baochen Sun. Bidirectional parallel feature pyramid network for object detection. *IEEE Access*, 10:49422–49432, 2022.

[Zhao and Wu 2019] Ting Zhao and Xiangqian Wu. *Pyramid Feature Attention Network for Saliency detection*, 2019.

[Zhou *et al.* 2021] Xiaoxiong Zhou, Qiuming Gong, Yongqiang Liu, and Lijun Yin. Automatic segmentation of TBM muck images via a deep-learning approach to estimate the size and shape of rock chips. *Automation in Construction*, 126:103685, 2021.