Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.13130
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Battle Models: Inception ResNet vs. Extreme Inception for Marine Fish Object Detection

Djarot Hindarto
Prodi Informatika, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional, Jakarta, Indonesia
Djarot.hindarto@civitas.unas.ac.id

Abstract: Within the domain of deep learning applied to computer vision, there exists a significant emphasis on the competition between two prominent models, namely Inception ResNet and Xception, particularly in the field of marine fish object detection. The present study conducted a comparative analysis of two advanced neural network architectures in order to assess their efficacy in the identification and localization of marine fish species in underwater images. The two models underwent a rigorous evaluation, utilizing their capabilities in feature extraction. The findings indicate a complex performance landscape, wherein Inception ResNet exhibits remarkable accuracy in identifying marine fish objects, while Xception demonstrates superior computational efficiency. The present study elucidates the inherent trade-off between precision and computational expenditure, offering valuable perspectives on the pragmatic ramifications of choosing one model over another. Furthermore, this research underscores the significance of carefully choosing a suitable model that aligns with the particular requirements of object detection applications in the context of marine fish. This study endeavors to guide professionals and scholars in marine biology and computer vision, enabling them to make well-informed choices when utilizing deep learning techniques to detect maritime fish objects in underwater settings. The research specifically focuses on the comparison between Inception ResNet and Xception models.

Keywords: Computer Vision; Detect Maritime Fish Objects; Deep Learning; Inception ResNet model; Xception models

INTRODUCTION

In the current epoch characterized by swift technological progressions and an increasing demand for advanced data analysis, computer vision has emerged as a highly captivating field with significant potential. Deep learning, employing progressively intricate and advanced neural network architectures (R. Zhang et al., 2023), has facilitated a more profound comprehension of image processing and object recognition within diverse applications, including marine biology. The identification and classification of marine fish pose significant challenges within the field of aquatic ecology, as they play a crucial role in monitoring and conserving the marine environment. The urgency to maintain the equilibrium of marine ecosystems has led to a growing demand for automated and precise techniques to detect maritime fish objects in underwater images.

This study aims to compare the performance of two prominent deep learning models, Inception ResNet (S. Yu et al., 2023) and Xception, in the context of a particular research domain. Both models possess a solid reputation in computer vision and have proven their ability to address complex object recognition objectives effectively. However, a pertinent question arises regarding the model's superior performance detecting marine fish objects. To answer this question, a thorough analysis will compare and contrast the two models to determine their respective benefits and drawbacks in this specific context. The size of bubbles and the velocity of foam in underwater images are crucial factors that have a significant impact on the effectiveness of fish object detection. The evaluation of flotation systems in industrial processes involving steel requires the evaluation of two key characteristics: bubble size and froth velocity. These parameters are instrumental in determining the quality and efficacy of such processes. Therefore, precise measurement tools for assessing bubble size and froth velocity in submerged imagery can significantly boost industrial efficiency and output.

Prior studies have employed diverse methodologies to quantify the dimensions of bubbles and the speed of froth movement in images captured underwater. Nevertheless, many of these methodologies depend on intricate and frequently laborious image processing techniques. The issue of overcoming the limitations inherent in conventional image processing algorithms has prompted the exploration of alternative approaches, such as deep learning. This method is appealing due to its advanced capabilities in feature extraction and classification. This study presents a novel methodology that employs a pre-existing convolutional neural network (L. Yu et al., 2022) to quantify the dimensions of bubbles and the velocity of froth in images captured underwater. The pre-trained convolutional neural network is a sophisticated deep learning model that has undergone training using extensive

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.13130
e-ISSN : 2541-2019
p-ISSN : 2541-044X

datasets, enabling it to identify intricate patterns within images effectively. Using pre-trained Convolutional Neural Networks (Sun et al., 2020) is anticipated to yield improved accuracy and efficiency in assessing bubble size and froth velocity in underwater images.

The approach, as mentioned above, exhibits significant promise across a range of applications, encompassing marine ecological monitoring and the steel industry. By employing precise and efficient techniques for measuring bubble size and froth velocity, we can enhance our comprehension of alterations occurring within the marine environment. Consequently, we can adopt appropriate measures to ensure the preservation of the sustainability of marine ecosystems. Conversely, within the steel industry, optimizing monitoring and control procedures about froths yields enhanced efficiency, resulting in time and resource savings. Before attaining this objective, it is imperative to conduct a meticulous evaluation of various deep learning models (Deepanshi et al., 2023), (Lee et al., 2023) to ascertain the selection of the most appropriate model for this particular undertaking. The relevance of the battle between Inception ResNet and Xception lies in its significance within the academic discourse. The present study will encompass a comprehensive examination of the models mentioned above, wherein an assessment will be conducted to determine the precision, computational efficacy, and efficacy in generating marine fish object detection.

Consequently, this study addresses fundamental questions in computer vision and marine biology. The outcomes of our research will have a significant impact on the field of marine ecological monitoring and the steel industry. In addition, our research will provide a deeper understanding of what deep learning models can do, which play a vital role in recognizing objects in underwater images. Do neural network architectures, such as Inception ResNet and Xception, have a substantial impact on the performance of object detection? (RQ 1). This inquiry pertains to the impact of various factors, namely the number of layers, kernel size, and transfer learning method, on the accuracy and efficiency of marine fish object detection when employing the Inception ResNet and Xception models. (RQ 2). The author employed these research inquiries to investigate the juxtaposition of Inception ResNet and Xception in the task of marine fish object detection while also discerning the factors that impact their performance within this particular context.

LITERATURE REVIEW

Previous scholarly literature has extensively examined Convolutional Neural Network (CNN) architectures, including Inception-ResNet and Exception, which have made noteworthy advancements in image recognition and object detection. The present study employs a Convolutional Neural Network (CNN) methodology to identify and classify multi-class micro facial expressions accurately. This is achieved through the utilization of a 2D-ResNet model. The findings indicate a significantly elevated level of precision, attaining a 99.3% rate. Additionally, the model demonstrated a recall rate of 99.12%, an F1 score of 0.98%, and a sensitivity of 99.16% (Durga & Rajesh, 2022). This reflects the model's efficacy in detecting micro facial expressions and its notable accuracy. Human Activity Recognition has been proposed for the ITH Gated Recurrent Unit (GRU). On various public datasets, the model achieves high F1 scores, reaching 96.27%, 90.05%, 90.30%, 99.12%, and 95.99%. GRU-INC combines GRU and Attention Mechanism for temporal aspects and Inception and Convolutional Block Attention Modules for spatial aspects (Rahman et al., 2023). Addresses the shortcomings of publicly available datasets that have yet to be curated for edge detection tasks. A new dataset is proposed that includes separate visual features for edges, contours, and boundaries. Unique architecture that can be trained from scratch without pre-trained weights (Soria et al., 2023). This technique comprises two submodules: Multiscale Signed RP (MSRP) and Inception Fully Convolutional Network (IFCN). MSRP addresses the shortcomings of RP, such as its inability to handle variations in scale and length of time series and its lack of clarity regarding trend direction. MSRP images are extracted from multiscale features using IFCN (Y. Zhang et al., 2022). This study intends to compare Inception-ResNet and Extreme Inception in the context of detecting objects in marine fish. The advantages and disadvantages of each model, as well as their potential contributions to the advancement of object detection technology, will be clarified by comparing these two models. By comparing the performance of the two, this research will provide a more thorough understanding of the selection of the most appropriate model for object detection applications in marine fish, as well as additional information for future research in this domain.

METHOD

Computer Vision

The academic discipline of artificial intelligence, commonly referred to as "computer vision," is dedicated to the examination and interpretation of visual information. The capacity of computational systems to comprehend and analyze visual data, particularly images and videos. This is an interdisciplinary field that combines mathematical principles, statistical methodologies, and signal-processing techniques. Its primary purpose is to develop algorithms and construct models that facilitate computer systems' visual perception and comprehension of their surroundings.

Object detection is a fundamental component of Computer Vision, which pertains to the ability of computers to discern and precisely determine the presence and location of objects within images or videos. The technology possesses many practical applications in various domains, such as security surveillance, autonomous vehicles, and

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.13130
e-ISSN : 2541-2019
p-ISSN : 2541-044X

medical processing. Furthermore, Computer Vision is employed in medical image processing to identify and examine patterns within medical images, including X-rays, MRIs, and CT scans, to diagnose diseases. Computer Vision is commonly employed in facial recognition, enabling computers to accurately identify individuals by analyzing their facial features. The technology possesses a broad spectrum of applications in security, identity management, and the development of personalized technologies, exemplified by the implementation of facial recognition on smartphones.

Object tracking, 3D reconstruction, gesture recognition, and image processing for character or handwriting recognition are additional computer vision applications. Computer Vision is essential for developing autonomous vehicles that enable vehicles to "see" and respond to their environment, including recognition of traffic signs, surrounding vehicles, and road mapping. In an industrial setting, Computer Vision is utilized for automated inspections of production lines, allowing for the detection of defects or changes in the manufactured products. This improves production quality and efficiency. Computer Vision enables computers to interact with the physical world using visual data. This technology's advancements have cultivated a vast array of relevant applications in various industries, bringing enormous benefits to monitoring, identification, automation, and visual data analysis.

Deep Learning

Deep Learning is a specialized field within artificial intelligence (Büyükşahin & Ertekin, 2019) that endeavors to create algorithms and models based on neural networks. The primary objective of this discipline is to enable the comprehension, manipulation, and extraction of intricate features from highly intricate data, particularly data that exhibit non-linear characteristics. The distinguishing characteristic of Deep Learning, compared to other methodologies, lies in the increased depth or quantity of layers within the neural network architecture. This depth facilitates the generation of progressively abstract and comprehensive data representations. One of the primary attributes of Deep Learning is its capacity to acquire knowledge from data. This implies that Deep Learning algorithms can adjust and enhance their performance in response to using additional data for training purposes. This attribute renders it highly efficient in various tasks, encompassing pattern recognition, classification, object detection, natural language processing, and other related applications.

Deep Learning is frequently linked to artificial neural networks, which draw inspiration from the architecture and operation of human neural networks. Deep Learning utilizes neural networks, which are composed of interconnected nodes or neurons arranged in layers. The input data undergoes a sequential process in which each layer is tasked with extracting features that are progressively more abstract and intricate. The widespread adoption of Deep Learning has resulted in many practical implementations. In computer vision, Deep Learning techniques are employed to perform tasks such as facial recognition in images, object detection in videos, and interpretation of medical imagery. In natural language processing, Deep Learning algorithms (Jahandad et al., 2019) enable computer systems to comprehend and generate text that closely resembles human language. Deep Learning is a prominent technique employed within the automotive industry to advance autonomous vehicles. Its application enables vehicles to identify and discern various elements, such as traffic signs, pedestrians, and other vehicles. The progress of deep Learning is inherently intertwined with advancements in computing technology and the accessibility of vast amounts of data. Deep Learning is rapidly advancing due to increasing computational capabilities and the availability of enormous amounts of data. Consequently, it has emerged as a highly promising and captivating domain within artificial intelligence.

Inception ResNet model

Inception-ResNet, is one of the neural network architecture models in the world of Deep Learning that has proven itself to be one of the most advanced in object recognition and image classification tasks. This model is the result of a combination of two well-known architectures, namely Inception and ResNet. Inception is known for its ability to extract features from various scales and levels of complexity in images, while ResNet is known for its residual block concept which allows network training with very deep layers without experiencing performance degradation problems. The main advantage of Inception ResNet is its ability to combine features from different levels of resolution and abstraction in an image. Thus, this model can recognize objects in images with a very high level of accuracy and is able to overcome the difficulty of training very deep networks. Inception ResNet is also known for its ability to overcome the vanishing gradient problem, which often occurs when training deep networks.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.13130
e-ISSN : 2541-2019
p-ISSN : 2541-044X

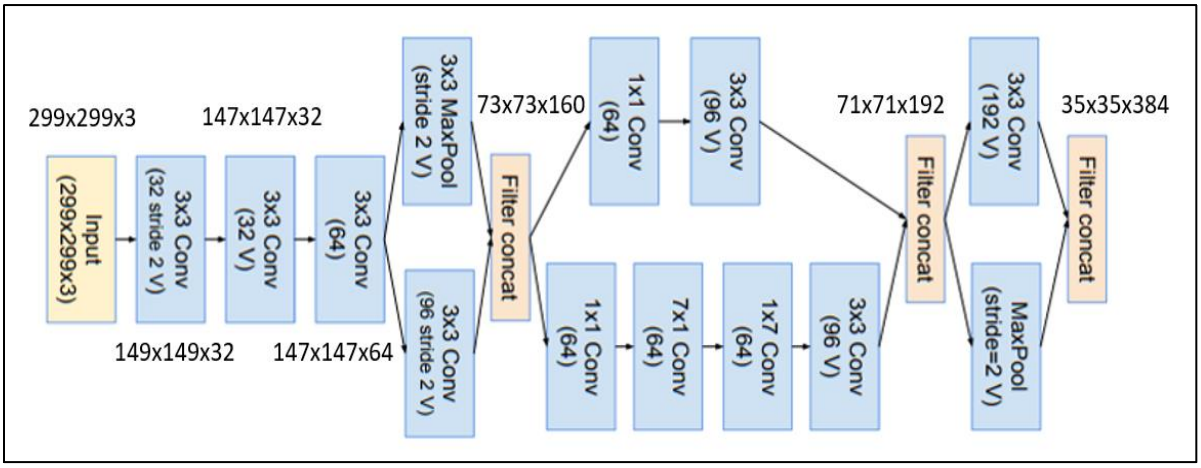

Figure 1. Inception-ResNet-v2 networks
Source: (Szegedy et al., n.d.)

Figure 1, The Inception ResNet architecture consists of a number of interconnected residual blocks, which allows smooth and efficient flow of information in the network. Each residual block consists of convolutional layers combined with Inception layers, creating a complex but highly efficient structure. The ResNet Inception model has been successfully applied in a variety of tasks, including image classification, object detection, image segmentation, and more. Its success in well-known image recognition competitions such as the ImageNet Competition has made this model very popular in the research and industrial communities.

However, although Inception ResNet has many advantages, its use requires considerable computing resources and a significant number of parameters. Therefore, implementation may require a powerful computing infrastructure. Despite these challenges, Inception ResNet remains one of the most powerful and reliable models in image processing and object recognition.

Xception models

Xception, also known as "Extreme Inception," is a convolutional neural network (Gawade et al., 2023) model that has garnered significant attention and acclaim in Deep Learning due to its revolutionary and highly sophisticated nature. The model under discussion was formulated by Google Research in the year 2017 and represents one of the most recent advancements in the field of convolutional neural network (Suherman et al., 2023) architecture. Xception is a novel convolutional neural network architecture that expands upon the established Inception concept, incorporating innovative advancements that revolutionize our comprehension and design of such networks. One of the primary advancements in the Xception model involves implementing a technique known as "depth wise separable convolution." The depth wise convolution technique divides the convolution process into two distinct stages. In the first stage, convolution is performed individually on each channel within the input data. Subsequently, in the second stage, convolution is conducted using a kernel that combines the results obtained from the initial step. The approach mentioned above demonstrates high efficiency by minimizing the parameter count necessary for convolutional operations while simultaneously preserving the network's capacity to extract intricate features.

The Xception architecture also introduces the novel concept of "factorized convolution," which involves decomposing spatial-based convolutions into separate spatial and channel-wise convolutions. This enables the network to concurrently process spatial and channel information, leading to enhanced and more effective feature representation. The primary benefit of Xception lies in its capacity to extract features at various levels of resolution and abstraction. This characteristic enables the network to achieve a remarkably accurate performance in object recognition tasks. The Xception model has demonstrated its effectiveness in various applications, encompassing tasks such as image classification, object detection, image segmentation, and even the advancement of autonomous vehicles. Despite the numerous advantages of Xception, its utilization necessitates substantial computational resources owing to its intricate and deep architecture.

Nevertheless, the outcomes achieved thus far have demonstrated the worth and potential of this model across diverse applications. The Xception model has emerged as a highly influential and effective tool in image processing and object recognition. It has gained significant prominence and remains widely employed in academic research and industrial applications to address complex visual tasks.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.13130

e-ISSN : 2541-2019
p-ISSN : 2541-044X

RESULT

Dataset

A valuable asset in developing object detection models is a dataset comprising 6300 images in the training dataset, 2700 images in the validation dataset, and 1890 images in the test dataset, encompassing nine distinct classes. The dataset presented herein offers a comprehensive range of images and classes that are essential for effectively training and evaluating object detection models, thereby ensuring high levels of accuracy. The model development is based on a training dataset comprising 6300 images. Within this dataset, the model can acquire knowledge about diverse features and attributes that are linked to each class. During the training process, a validation dataset comprising 2700 images was employed to assess the model's performance. This evaluation aimed to ascertain that the model not only memorized the training data but also demonstrated the ability to generalize its knowledge by accurately recognizing objects that were previously unseen.

The final test dataset, which comprises 1890 images, is utilized to evaluate the accuracy and performance of the object detection model. During this phase, the model will undergo testing using images that have not been previously encountered. The outcomes of this examination will yield insights into the model's capacity to discern objects within authentic scenarios accurately. Datasets that exhibit variations in the number of classes present pose a realistic challenge in the context of object detection tasks. Models must possess the capability to discern between various categories of objects. The dataset in question provides an appropriate structure for training models that exhibit a high degree of accuracy in object recognition. The outcomes of this study, based on a dataset that accurately represents commonplace scenarios, will hold substantial significance in advancing the progress of more advanced object detection methodologies.

Inception ResNet model

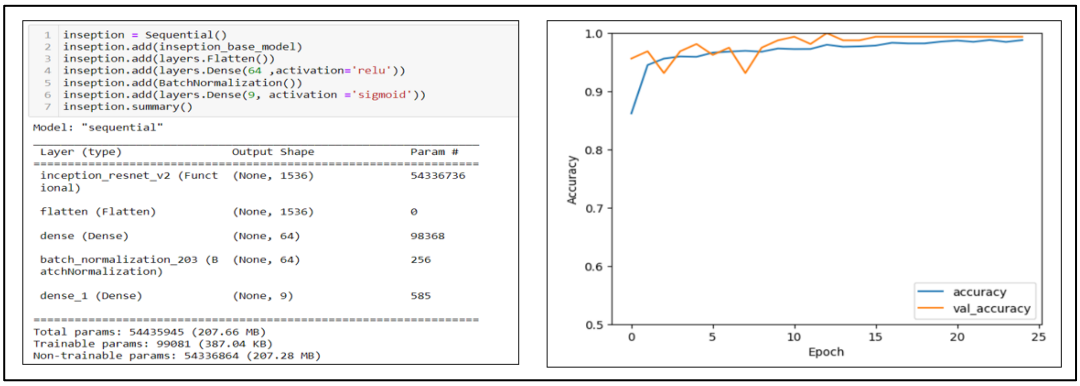

Figure 2. Architecture Inception-ResNet-v2 in Python
Source: Researcher Property

The Inception-ResNet-v2 is a convolutional neural network architecture that amalgamates fundamental characteristics of the Inception and ResNet models. The model under consideration is a product of Google's development efforts. It integrates the diverse filter sizes of Inception with the residual features of ResNet, thereby yielding a deep model characterized by its notable efficiency. The Inception-ResNet-v2 model is widely recognized for its proficiency in extracting features across multiple scales. It is frequently employed in various applications, including image classification and object detection. This architecture has been used utilizing the Python programming language and deep learning libraries like TensorFlow or Keras, rendering it a valuable asset in the advancement of intricate and precise models in the field of image processing.

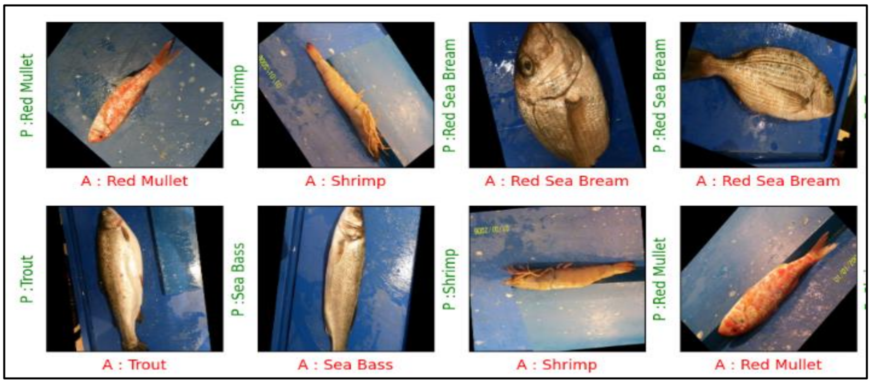

Figure 3. Inception-ResNet-v2 algorithm in the identification of fish species
Source: Researcher Property

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.13130

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Figure 3, The application of the Inception-ResNet-v2 algorithm in the identification of fish species, such as Red Mullet, Shrimp, Red Sea Bream, Trout, and Sea Bass, has yielded highly favorable outcomes. The algorithm exhibits exceptional accuracy, demonstrating a remarkable ability to identify various types of fish with a significant level of precision. The benefits mentioned above possess significant value across diverse contexts, particularly within the domains of fisheries and conservation. The utilization of the Inception-ResNet-v2 algorithm yields outcomes that are both precise and accurate, thereby contributing to the effective monitoring of fish populations, safeguarding endangered species, and promoting sustainable management of fisheries. This technology facilitates the categorization of fish in visual media, thereby facilitating the creation of automated systems for monitoring fish presence and behavior in marine ecosystems. This process constitutes a crucial stage in comprehending and conserving subaquatic biodiversity.

Xception models

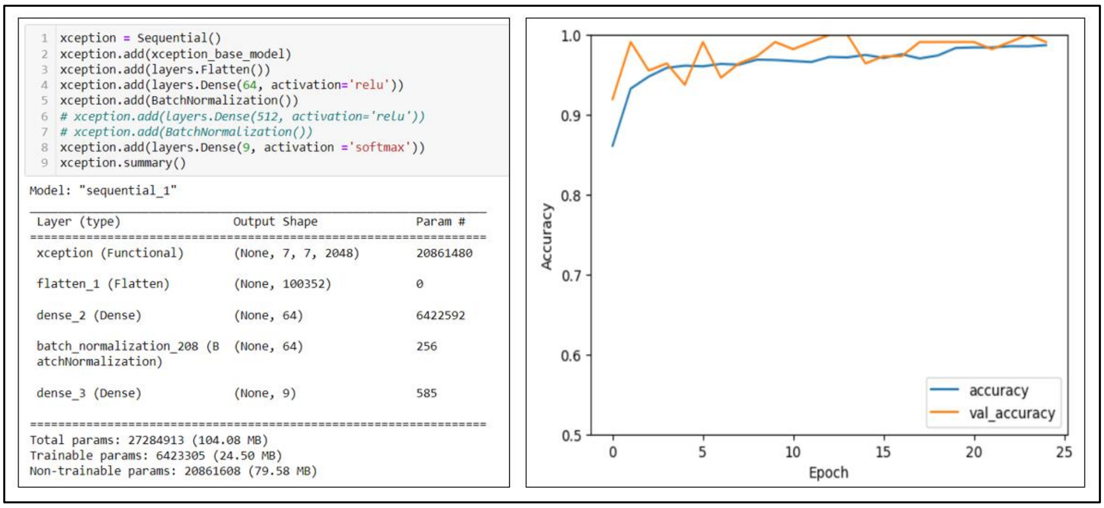

Figure 4. Architecture Extreme Inception (Exception) in Python
Source: Researcher Property

Figure 4, Extreme Inception, commonly called Exception, is a neural network framework that amalgamates distinctive components from Inception and other architectures, resulting in a robust model specifically designed for image recognition. The Exception algorithm has been specifically developed to extract features at various scales, thereby facilitating the accurate detection of objects. The utilization of the "depth wise separable convolution" technique in Exception results in a reduction of the model parameters, thereby enhancing its computational efficiency. The utilization of exceptions in image recognition applications, particularly in object detection, has gained significant popularity with the adoption of the Python programming language and its associated deep learning libraries, such as TensorFlow or Keras. The exceptional capabilities of Exception in feature extraction and recognition accuracy render it an indispensable tool across diverse domains such as computer vision, image processing, and visual content comprehension.

Figure 5. Exception algorithm in the identification of fish species
Source: Researcher Property

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.13130
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Figure 5 depicts the utilization of the Exception algorithm in identifying fish species. The algorithm mentioned above has demonstrated efficacy in the domain of image processing, specifically in the task of discerning distinct fish species. Exception enables us to accurately and precisely classify fish by leveraging its remarkable capacity to determine the distinguishing characteristics exhibited by each species. The algorithm's identification results have proven valuable in diverse contexts, including fisheries monitoring and conservation efforts. Within marine science, acquiring precise knowledge about fish species plays a pivotal role in effectively managing marine resources with a focus on sustainability. Figure 5 illustrates the significant contribution of Exception in this endeavor, as it offers dependable identification outcomes and empowers users to comprehend and sustain biodiversity within marine ecosystems.

DISCUSSIONS

Do neural network architectures, such as Inception ResNet and Xception, have a substantial impact on the performance of object detection? (RQ 1).

The neural network architecture plays a crucial role in determining the capacity of a model to comprehend and interpret visual information derived from image data. Neural network architectures possess distinct characteristics that set them apart from one another. ResNet and Xception, two prominent architectures in deep learning, present particular methodologies for structuring layers and operations within a neural network. ResNet architecture integrates the principles of Inception with residue blocks, enabling the model to combine features from various levels of abstraction efficiently. In contrast, Xception introduces the notion of "depth wise separable convolution," a technique that effectively decreases the parameter count and enhances the efficiency of information processing.

It is crucial to remember that every object detection task can exhibit distinct characteristics, including variations in lighting conditions, backgrounds, and object sizes that necessitate identification. Hence, the selection of an appropriate neural network architecture can significantly influence the performance of object detection. An architectural design that demonstrates efficiency in extracting features across various scales and abstractions has the potential to yield superior outcomes when addressing variations in object detection tasks. Furthermore, network performance can be influenced by multiple factors, including the number of layers, kernel size, and the method employed for transfer learning. The accuracy and efficiency of object detection can be enhanced by determining the optimal settings for these parameters. Hence, conducting an extensive investigation to address this inquiry necessitates meticulous experimentation and statistical evaluation to compare the efficacy of Inception ResNet and Xception in diverse scenarios of marine fish object detection tasks.

The findings of this study will offer significant understanding regarding the impact of neural network architecture on the performance of object detection, as well as the most effective configuration for this particular task. The provided information has the potential to enhance the advancement of more effective and proficient object detection models within this specific context, as well as in other domains that necessitate image analysis and object recognition.

This inquiry pertains to the impact of various factors, namely the number of layers, kernel size, and transfer learning method, on the accuracy and efficiency of marine fish object detection when employing the Inception ResNet and Xception models. (RQ 2).

The number of layers within a neural network architecture can impact the network's capacity to extract distinctive characteristics and effectively represent connections within image data. The model's ability to recognize more abstract traits gets better as the depth of a neural network grows. However, this in-depth augmentation can concurrently lead to an escalation in computational complexity. The impact of the number of layers will primarily rely on the intricacy of the marine fish object detection assignment, as specific tasks may necessitate deeper networks to handle the more intricate variations present in the underwater environment. The resolution and coverage of features extracted by the network can be influenced by the kernel size used in convolution operations. A larger kernel size can facilitate the network in discerning more expansive features, whereas a smaller kernel size can aid in detecting more intricate features. The optimal configurations will be contingent upon the specific attributes of the task at hand and the dataset employed for detecting marine fish objects.

Furthermore, using transfer learning techniques holds significant importance in advancing object detection models. By leveraging the acquired knowledge from prior models via transfer learning, the model can expedite convergence and enhance its accuracy. Nevertheless, selecting an appropriate transfer learning approach can significantly influence the ultimate performance of the model. Conducting research that incorporates experimental manipulation of layer number, kernel size, and transfer learning techniques will contribute to a more comprehensive comprehension of the impact of these variables on the accuracy and efficiency of marine fish object detection. The findings of this study have the potential to offer valuable insights for advancing object detection models that are more efficient and precise in intricate marine settings.

By acquiring a more comprehensive comprehension of the interplay among these variables, scholars and professionals in object detection can enhance the configuration of their models to detect marine fish objects. Furthermore, this knowledge can be effectively employed in image processing and object identification tasks beyond the scope of marine fish detection.

e-ISSN : 2541-2019
p-ISSN : 2541-044X

CONCLUSION

This study conducted a comparative analysis of two neural network architectures, specifically Inception ResNet and Xception, within the domain of marine fish object detection. The experimental findings demonstrate that both entities possess robust capabilities in identifying objects within diverse underwater environmental contexts, exhibiting a notable degree of precision. Nevertheless, a noteworthy observation arises when contemplating computational efficiency. Xception, through its utilization of the "depth wise separable convolution" concept, demonstrates its efficacy as a more efficient option regarding parameter count and computational operations, thereby enhancing the speed of training and inference procedures. While Inception ResNet shows strong performance, particularly in the presence of variations in lighting and background, Xception emerges as a promising alternative for rapid and efficient detection of underwater objects. The findings of this study offer significant insights for developers of object detection solutions in intricate marine environments.

Furthermore, this study also investigates the influence of various factors, including the number of layers, kernel size, and transfer learning techniques, on the performance of both architectures. It aims to identify the most effective configurations and facilitate the creation of more efficient models for this particular task. In summary, when deciding between Inception ResNet and Xception to detect marine fish objects, it is crucial to consider the efficiency and accuracy aspects, as they play a significant role in meeting the specific task requirements.

REFERENCES

Büyükşahin, Ü. Ç., & Ertekin, Ş. (2019). Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. *Neurocomputing*, *361*, 151–163. https://doi.org/10.1016/j.neucom.2019.05.099

Deepanshi, Budhiraja, I., Garg, D., & Kumar, N. (2023). Choquet integral based deep learning model for COVID-19 diagnosis using eXplainable AI for NG-IoT models. *Computer Communications*, *212*(September), 227–238. https://doi.org/10.1016/j.comcom.2023.09.032

Durga, B. K., & Rajesh, V. (2022). A ResNet deep learning based facial recognition design for future multimedia applications. *Computers and Electrical Engineering*, *104*(PA), 108384. https://doi.org/10.1016/j.compeleceng.2022.108384

Gawade, S., Bhansali, A., Patil, K., & Shaikh, D. (2023). Application of the convolutional neural networks and supervised deep-learning methods for osteosarcoma bone cancer detection. *Healthcare Analytics*, *3*(February), 100153. https://doi.org/10.1016/j.health.2023.100153

Jahandad, Sam, S. M., Kamardin, K., Amir Sjarif, N. N., & Mohamed, N. (2019). Offline signature verification using deep learning convolutional Neural network (CNN) architectures GoogLeNet inception-v1 and inception-v3. *Procedia Computer Science*, *161*, 475–483. https://doi.org/10.1016/j.procs.2019.11.147

Lee, S., Jeong, H., Hong, S. M., Yun, D., Lee, J., Kim, E., & Cho, K. H. (2023). Automatic classification of microplastics and natural organic matter mixtures using a deep learning model. *Water Research*, *246*(September). https://doi.org/10.1016/j.watres.2023.120710

Rahman, T., Amatullah, M., Afreen, S., & Abu, M. (2023). *GRU-INC : An inception-attention based approach using GRU for human activity recognition*. *216*(November 2022).

Soria, X., Sappa, A., Humanante, P., & Akbarinia, A. (2023). Dense extreme inception network for edge detection. *Pattern Recognition*, *139*. https://doi.org/10.1016/j.patcog.2023.109461

Suherman, E., Hindarto, D., Makmur, A., & Santoso, H. (2023). Comparison of Convolutional Neural Network and Artificial Neural Network for Rice Detection. *Sinkron*, *8*(1), 247–255. https://doi.org/10.33395/sinkron.v8i1.11944

Sun, P., Liu, P., Li, Q., Liu, C., Lu, X., Hao, R., & Chen, J. (2020). DL-IDS: Extracting Features Using CNN-LSTM Hybrid Network for Intrusion Detection System. *Security and Communication Networks*, *2020*, 8890306. https://doi.org/10.1155/2020/8890306

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (n.d.). *the Impact of Residual Connections on Learning*.

Yu, L., Xue, L., Liu, F., Li, Y., Jing, R., & Luo, J. (2022). The applications of deep learning algorithms on in silico druggable proteins identification. *Journal of Advanced Research*, *xxxx*. https://doi.org/10.1016/j.jare.2022.01.009

Yu, S., Xie, L., & Huang, Q. (2023). Inception convolutional vision transformers for plant disease identification. *Internet of Things (Netherlands)*, *21*(288), 1–18. https://doi.org/10.1016/j.iot.2022.100650

Zhang, R., Gao, M. R., Zhang, P. Y., Zhang, Y. M., Fu, L. H., & Chai, Y. F. (2023). Research on an ultrasonic detection method for weld defects based on neural network architecture search. *Measurement: Journal of the International Measurement Confederation*, *221*(August), 113483. https://doi.org/10.1016/j.measurement.2023.113483

Zhang, Y., Hou, Y., OuYang, K., & Zhou, S. (2022). Multi-scale signed recurrence plot based time series classification using inception architectural networks. *Pattern Recognition*, *123*. https://doi.org/10.1016/j.patcog.2021.108385