JACOBS
UNIVERSITY

Deep Learning for Detecting Amphoras in Ancient Shipwrecks

by

Tianyao Chen

Bachelor Thesis in Computer Science

Submission: April 15, 2021 Supervisor: Prof. Dr. Andreas Birk

---

Jacobs University Bremen | Department of Computer Science and Electrical Engineering

English: Declaration of Authorship

I hereby declare that the thesis submitted was created and written solely by myself without any external support. Any sources, direct or indirect, are marked as such. I am aware of the fact that the contents of the thesis in digital form may be revised with regard to usage of unauthorized aid as well as whether the whole or parts of it may be identified as plagiarism. I do agree my work to be entered into a database for it to be compared with existing sources, where it will remain in order to enable further comparisons with future theses. This does not grant any rights of reproduction and usage, however.

This document was neither presented to any other examination board nor has it been published.

German: Erklärung der Autorenschaft (Urheberschaft)

Ich erkläre hiermit, dass die vorliegende Arbeit ohne fremde Hilfe ausschließlich von mir erstellt und geschrieben worden ist. Jedwede verwendeten Quellen, direkter oder indirekter Art, sind als solche kenntlich gemacht worden. Mir ist die Tatsache bewusst, dass der Inhalt der Thesis in digitaler Form geprüft werden kann im Hinblick darauf, ob es sich ganz oder in Teilen um ein Plagiat handelt. Ich bin damit einverstanden, dass meine Arbeit in einer Datenbank eingegeben werden kann, um mit bereits bestehenden Quellen verglichen zu werden und dort auch verbleibt, um mit zukünftigen Arbeiten verglichen werden zu können. Dies berechtigt jedoch nicht zur Verwendung oder Vervielfältigung.

Diese Arbeit wurde noch keiner anderen Prüfungsbehörde vorgelegt noch wurde sie bisher veröffentlicht.

Date, Signature

Abstract

Consider this a separate document, although it is submitted together with the rest. The abstract aims at another audience than the rest of the proposal. It is directed at the final decision maker or generalist, who typically is not an expert at all in your field, but more a manager kind of person. Thus, don't go into any technical description in the abstract, but use it to motivate the work and to highlight the importance of your project.

(target size: 15-20 lines)

Contents

1 Introduction **1**
- 1.1 Motivation . . . 1
  - 1.1.1 Relevance of Amphoras . . . 1
  - 1.1.2 Computer Vision for Underwater Archaeology . . . 2
- 1.2 Deep Learning . . . 3
  - 1.2.1 Artifical Neural Networks (ANNs) . . . 3
  - 1.2.2 Convolutional Neural Networks (CNNs) . . . 5
- 1.3 Deep Learning vs. Traditional Computer Vision . . . 7
- 1.4 Object Detection . . . 8
  - 1.4.1 Fully Convolutional Networks (FCNs) . . . 9
  - 1.4.2 General Object Detection Framework Components . . . 9
  - 1.4.3 Region-Based Convultional Neural Networks (R-CNN) . . . 9
  - 1.4.4 Single Shot Detector (SSD) . . . 9
  - 1.4.5 You Only Look Once (YOLO) . . . 9

2 Related Work **10**

3 Data and Methods **10**
- 3.1 Data . . . 10
- 3.2 Model . . . 10
- 3.3 Model Training . . . 10

4 Evaluation **10**
- 4.1 Visual Evaluation . . . 11
- 4.2 Metric Evaluation . . . 11

5 Conclusions **11**

6 Future Work **11**

1 Introduction

1.1 Motivation

1.1.1 Relevance of Amphoras

The name *amphora* is derived from the Greek word *amphoreus*, which literally means "two-handled" [1, 2]. It is the combination of two linguistic roots: *amphi* (on both sides) and *phoreus* (bearer) [1, 2]. Amphoras (or amphorae) were commercially used from 1500 B.C.E. to 500 C.E. to ship products throughout the Mediterranean, supplying the ancient Greek and Roman empires [2]. Amphoras were designed to ship large quantities of liquid (wine, olives, and oils) and dry products (grain, nuts, and salted fish) [2].

Like many measures that are named after the packages, amphoras were also a semi-standard unit of liquid measure [2]. A cargo ship's capacity was measured by the number of amphoras it could carry instead of by weight [2, 3].

The structurally strong egg-like shape and the high volume-to-weight ratio made amphoras very efficient packages [2]. Amphoras were by far the most common cargo type in Mediterranean shipwreck analysis; more than half of the ships only carried amphoras [2, 4].

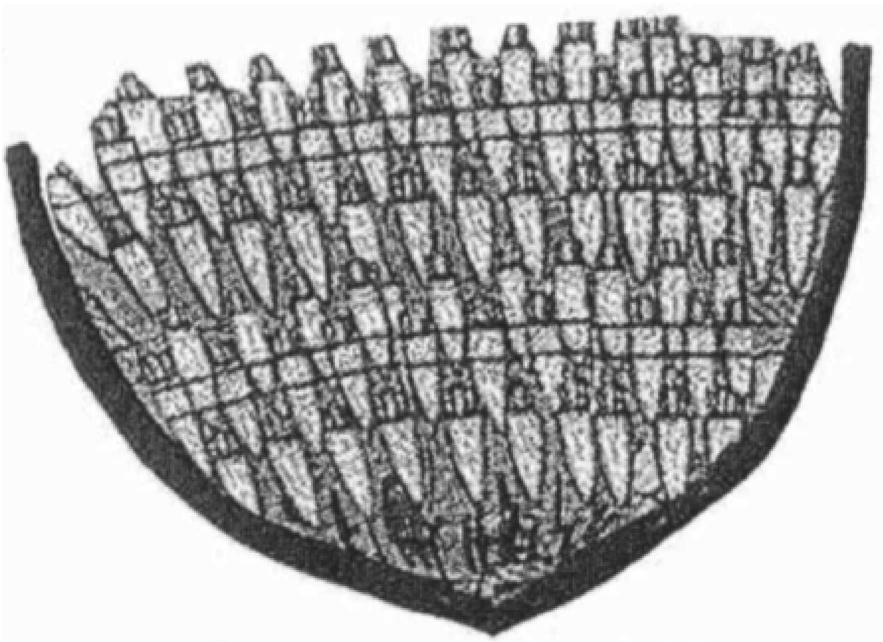

Figure 1: The egg-like shape enabled amphoras to interlock and minimize the waste of space on a ship. Source: [2].

Amphoras' various shapes and markings - which changed by time, region, producer, contents, and brand identity - were used to identify the package status and the different products inside [2].

Amphoras have great significance in archaeology. They can be used as evidence for the trade patterns throughout the Mediterranean [2]. As they were usually discarded at the destination of a trade and have been found in shipwrecks, archaeologists have been

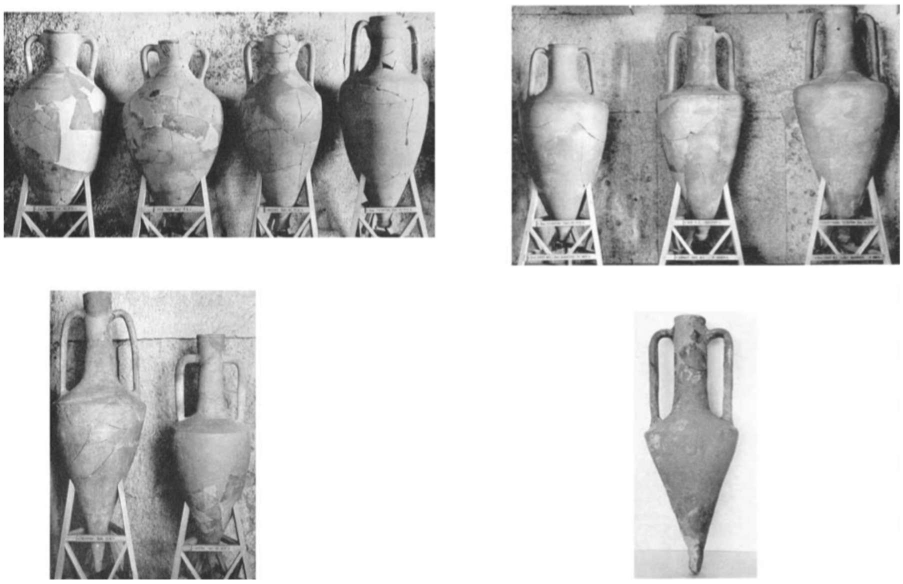

Figure 2: Amphoras have various shapes. Source: [2].

using them to recreate the transit routes [2]. Furthermore, researchers have been able to classify different amphoras, which also helps to date ruins and shipwrecks [2].

1.1.2 Computer Vision for Underwater Archaeology

Computer vision is the science of perceiving and understanding the world through images and videos [5]. There have been multiple exciting applications of computer vision, including image classification [6], object detection and localization [7, 8], art generation (neural style transfer) [9], image creation with Generative Artificial Networks (GAN) [10], face recognition [11], action and activity recognition [12], human pose estimation [13], and image recommendation system [14].

However, there is still limited research for the application of computer vision and machine learning in archaeology, especially underwater archaeology, compared to other domains [15, 16]. Computer vision, instead of visual inspection, could be used to automate the detection, assessment, and classification of artifacts [15].

Underwater computer vision has proven to be challenging, largely due to: 1) the distortion and attenuation caused by light propagation in water, and 2) the unrestricted natural environment with the abundance of marine life and suspended particles [16, 17, 18].

Despite the challenges, computer vision has lower cost [17] compared to sonar imagery [19] and laser scanning [20]. Plus, the increasingly abundant visual data obtained through autonomous underwater vehicles (AUVs), unmanned underwater vehicles (UUVs) [18, 21], and seafloor cabled observatories [16] enables us to utilize deep learning.

Furthermore, the research for deep-water shipwrecks is even more limited, mostly due to the lack of information and accessibility [22]. However, the need to study deep-water sites are in high demand, as the threats to these sites are increasing [22]. One major threat is the new forms of trawling that destroy the surface of these sites and interfere with the readability [22]. This means that many shipwrecks are likely to be damaged before they can be studied [22]. It is thus crucial to implement efficient, accessible, and accurate techniques like deep learning based computer vision to study deep-water shipwrecks.

1.2 Deep Learning

Machine learning is the class of algorithms that allow computers to learn and improve from data instead of being explicitly programmed [23, 24]. And deep learning is the subfield of machine learning that builds artificial neural networks with more than one layer between the input and output layers [24, 25, 26]. Deep learning constructs complex representations by combining simpler ones from the previous layers [27].

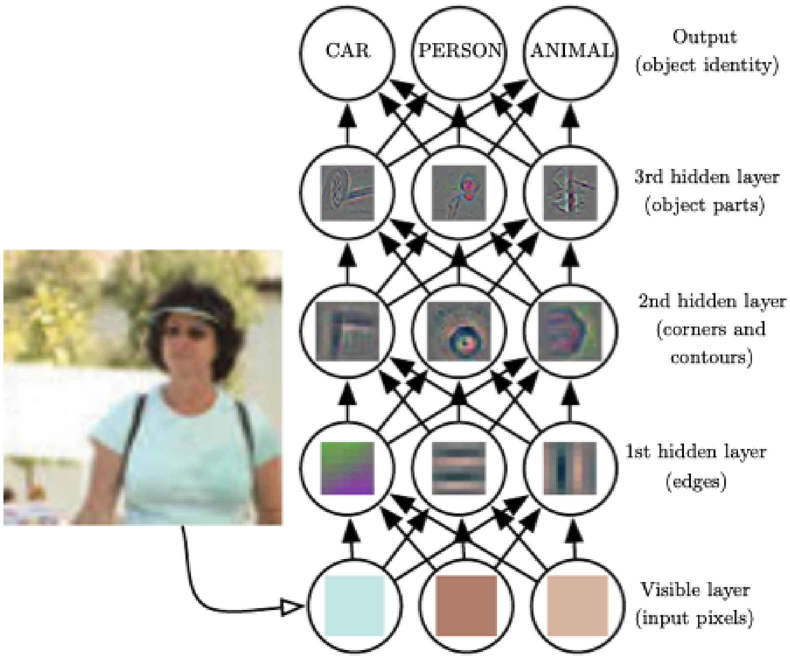

Figure 3: A deep learning system that learns the representations of a person. This is achieved by combining simpler features like corners and contours, which are further expressed by combining simpler features like edges. Source: [27].

1.2.1 Artifical Neural Networks (ANNs)

Inspired by the biological neuron, artificial neural networks (ANNs) were first introduced in 1943 using propositional logic [28]. The artificial neuron activates its single binary output when the number of active binary inputs reaches the activation threshold, which enables us to build networks that can perform any logical computation [24, 28].

Then the Perceptron was introduced in 1957, which is based on a different artificial neuron called threshold logic unit (TLU) or linear threshold unit (LTU) [29]. The inputs and outputs are numbers instead of binary values, and each input has a weight. TLU computes the weighted sum of the inputs and then applies a step function like the Heaviside

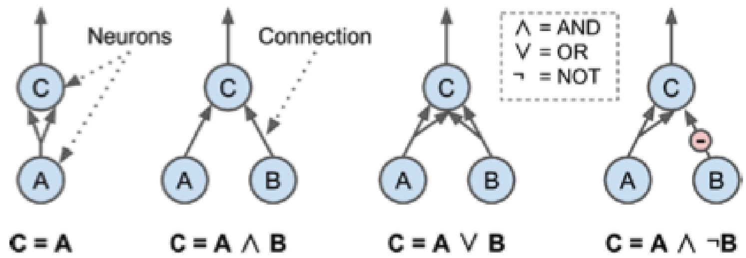

Figure 4: ANNs performing logical computations with the activation threshold of 2. Source: [24].

function $heaviside(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x \geq 0 \end{cases}$ [24, 29].

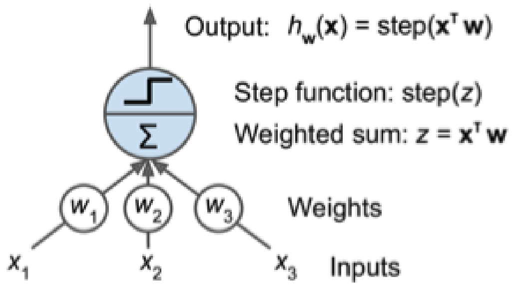

Figure 5: Threshold logic unit. Source: [24].

A single TLU can be used for simple linear binary classification, while a layer of TLUs plus a bias neuron form a Perceptron capable of multi-output classification [24].

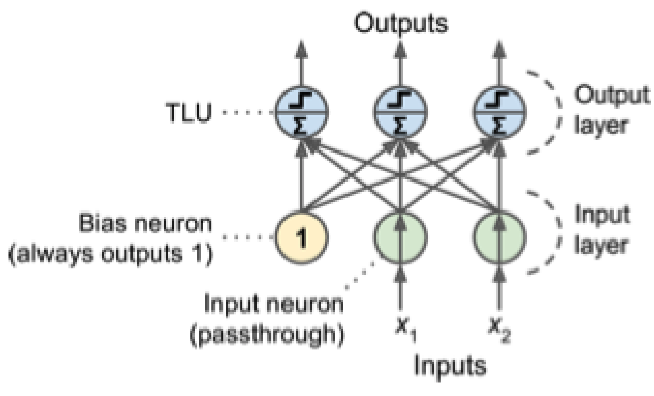

Figure 6: A Perceptron with three output neurons. Source: [24].

The outputs of a fully connected layer is computed as follows, where $\mathbf{X}, \mathbf{W}, \mathbf{b}$, and ϕ are respectively the input matrix, weight matrix, bias vector, and activation function:

$$h_{\mathbf{W},\mathbf{b}}(\mathbf{X}) = \phi(\mathbf{X}\mathbf{W} + \mathbf{b})$$

The Perceptron is trained using a variant of the Herbb's rule [30], which is famously summarized as "neurons wire together if they fire together [31]" [24]. However, the Perceptron can not learn complex patterns due the linear decision boundary of the output neurons, and it can only make predictions based on a hard threshold instead of outputting a class probability [24]. To address these limitations, the Multilayer Perceptron (MLP) was introduced by stacking multiple Perceptrons.

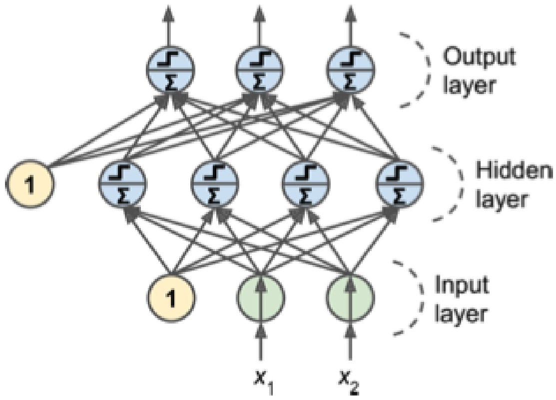

Figure 7: A Multilayer Perceptron with one hidden layer. Source: [24].

Backpropagation [32] is used to train the MLP, which first makes a prediction and measures the error in the forward pass, then measures the error contribution from each connection in the reverse pass, and finally tweaks the connection weights to reduce teh error in the Gradient Descent [33] step [24]. Activation functions like Rectified Linear Unit $ReLU(x) = max(0, x)$ are used to add nonlinearity, which theoretically gives a large enough deep neural network the ability to approximate any continuos function [24].

1.2.2 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) were inspired by the brain's visual cortex, and they have been used in computer vision since the 1980s [24]. We can not simply use a deep neural network with fully connected layers for computer vision, as it breaks down for large images due to the huge number of parameters it requires [24]. CNNs also have successful applications in other domains like recommender systems [34] and natural language processing (NLP) [35].

Hubel et al. [36, 37, 38] found that many biological neurons have a small receptive field, which means they only react to visual stimuli in a limited region of the visual field [24]. Some neurons only react to horizontal lines, while others only react to lines with different orientations [24]. Some neurons have larger receptive fields, and they react to more complex patterns formed by lower-level patterns [24].

Neurons in the first convolutional layer are only connected to pixels in their receptive fields, and neurons in the second convolutional layer are only connected to the neurons in a small receptive field in the first layer [24]. This allows the CNN to concentrate on lower-level features in the first hidden layer, then combine them into higher-level features in the second hidden layer, and so on [24].

The filters or convolution kernels, which are learned during training, are neurons' weights that can be presented as small images the size of receptive fields [24]. For example, a black sqaure with a horizontal white line in the middle (a matrix full of 0s except for the

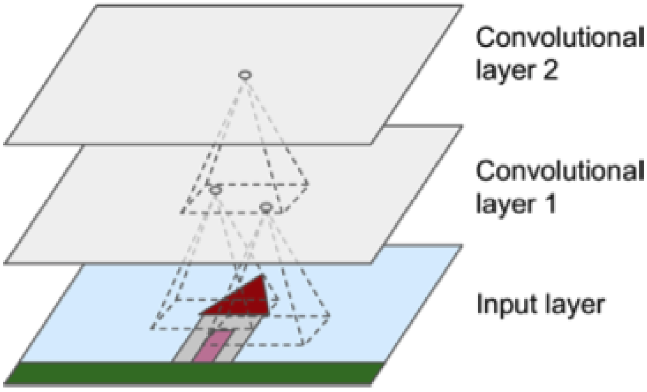

Figure 8: CNN layers with rectangular receptive fields. Source: [24].

central row with 1s) is a filter that only reacts to the central row in the receptive field. A layer of neurons with the same filter outputs a feature map, which highlights the parts of an image that activate the filter the most [24].

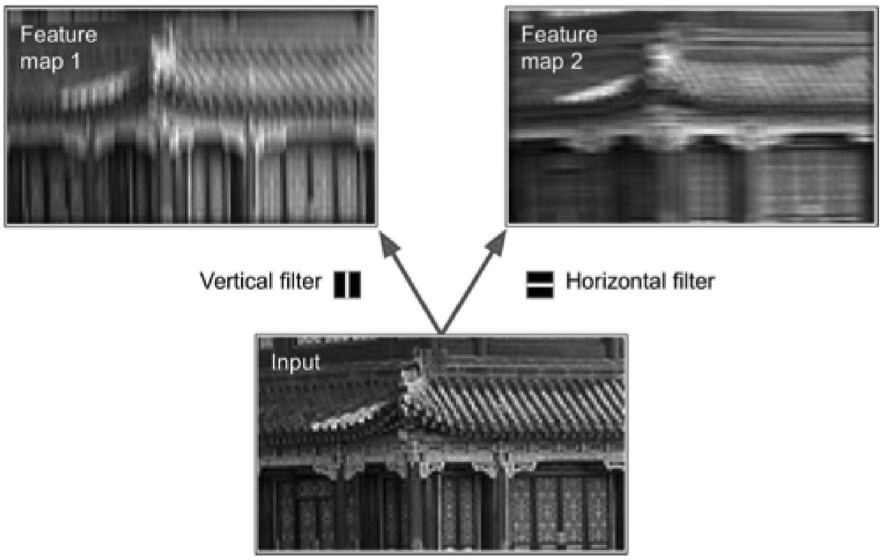

Figure 9: Two feature maps obtained by applying two different filters. Source: [24].

The pooling layers subsample (i.e. shrink) the input image to reduce the computational load and the number of paramters, which also reduces overfitting [24]. Plus, pool layers can bring some invariance to small translations, rotations, and scaling [24].

The typical CNN architecture involves the aforementioned convolutional layers, pooling layers, and fully connected layers. Some well-established CNN architecutres are LeNet-5 [39], AlexNet [40], GoogLeNet [41], VGGNet [42], ResNet (Residual Network) [43], Xception (Extreme Inception) [44], and SENet (Squeeze-and-Excitation Network) [45].

Here we give ResNet a more in-depth introduction, as it's the backbone network in the model we use to detect amphoras. ResNet is a fully convolutional network (FCN) [43, 46]. FCNs contain only convolutional layers and pooling layers, which means input images can have arbitrary sizes [46]. Resnet adds skip connections (or shortcut connections),

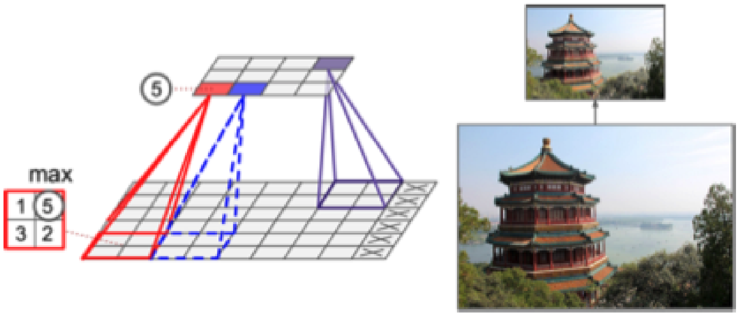

Figure 10: A max pooling layer with a 2 x 2 kernel and stride 2. Only the max value from each receptive field gets passed to the next layer. Source: [24].

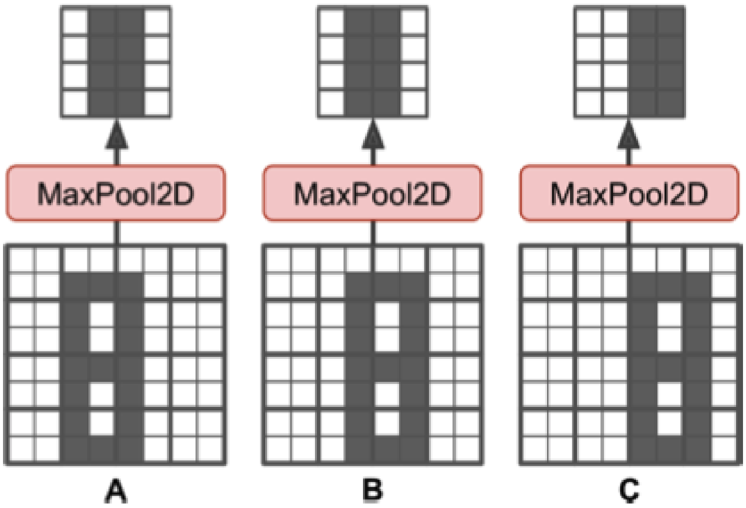

Figure 11: Max pooling layer's invariance to small translations. Source: [24].

which means the input signal of a layer is also added to the ouput of a higher up layer [24, 43]. Skip connections help speed up the training considerably, since: 1) the network preconditions the problem to be the identify function, which is often close to the target function, and 2) the network can start making progress even if some layers have not started learning yet [24, 43].

1.3 Deep Learning vs. Traditional Computer Vision

The increasingly abundant visual data and the improvements of deep learning algorithms, computing power, and image resolution have enabled deep learning to achieve state-of-the-art performance in many computer vision tasks, including image classification, object detection, and semantic segmentation [16, 47, 48].

Traditionally, computer vision requires the manual feature selection and engineering step, which relies on domain knowledge to produce high-quality features [5, 48]. These hand-crafted features are further procesed by a machine learning classifier like a support vector machine (SVM) [5, 48]. However, manual feature extraction becomes more and more complex as the number of classes increases [48]. Besides, conventional machine learning algorithms' ability to generalize saturates quickly as the size of training data grows

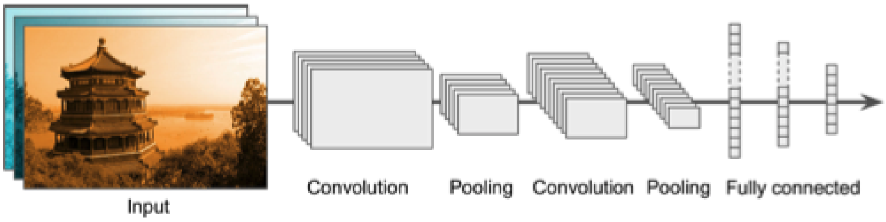

Figure 12: The typical CNN architecuture. Source: [24].

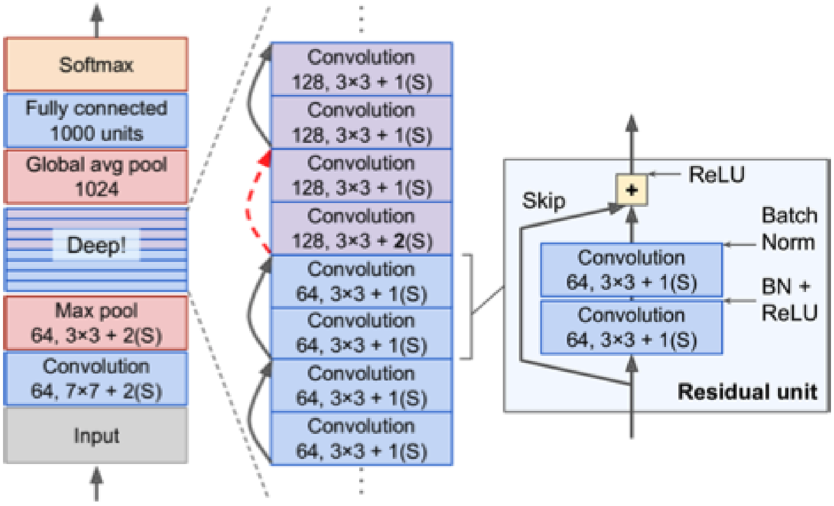

Figure 13: ResNet architecture. Source: [24].

[16].

In deep learning, the manual feature extraction step is no longer needed, and neural networks can be trained end-to-end as a feature extractor plus a classifier [5, 48]. Thus, deep learning requires less domain knowledge, it and provides more flexibility as the models can be re-trained with a custom dataset for any specific use case [48].

However, deep learning does not make traditional computer vision obsolete. Techniques like Scale Invariant Feature Transform (SIFT) [49], Speeded Up Robust Features (SURF) [50], and Features from Accelerated Segment Test (FAST) [51] are still useful in improving performance for computer vision tasks [48]. And deep learning's performance depends on obtaining large datasets with high image resolution [48]. Some popular public datasets like PASCAL Visual Object Classes (VOC) [52], ImageNet [53], and Microsoft Common OBjects in Context (COCO) [54] have respectively 500 thousand, 14 million, and 328 thousand images.

1.4 Object Detection

Define object detection and introduce the sliding CNN approach.

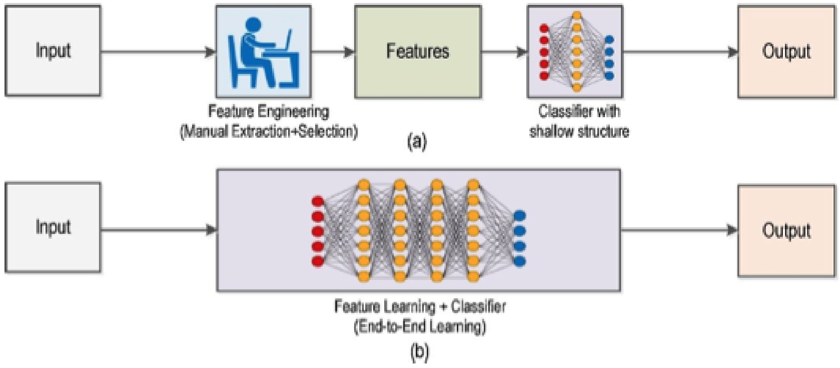

Figure 14: (a) Traditional computer vision workflow. (b) Deep learning workflow. Source: [48].

1.4.1 Fully Convolutional Networks (FCNs)

1.4.2 General Object Detection Framework Components

Region Proposals

Network Predictions

Non-Maximum Suppression (NMS)

Metrics

1.4.3 Region-Based Convultional Neural Networks (R-CNN)

R-CNN

Faster R-CNN

Faster R-CNN

1.4.4 Single Shot Detector (SSD)

1.4.5 You Only Look Once (YOLO)

YOLO

YOLOv2

YOLOv3

YOLOv4

YOLOv5

2 Related Work

3 Data and Methods

This is the technical core of the thesis. Here you lay out your how you answered your research question, you specify your design of experiments or simulations, point out difficulties that you encountered, etc.

(target size: 5-10 pages)

3.1 Data

3.2 Model

3.3 Model Training

4 Evaluation

This section discusses criteria that are used to evaluate the research results. Make sure your results can be used to published research results, i.e., to the already known state-of-the-art.

(target size: 5-10 pages)

Number	Description
7	A lucky number in Western culture
8	A lucky number in Chinese and other Asian cultures
42	Answer to the ultimate question of life, the universe, and everything
404	Not found

Table 1: Useless insights I gained with no further meaning

4.1 Visual Evaluation

4.2 Metric Evaluation

5 Conclusions

Summarize the main aspects and results of the research project. Provide an answer to the research questions stated earlier.

(target size: 1/2 page)

6 Future Work

References

[1] Douglas Harper et al. “Online etymology dictionary”. In: (2001).

[2] Diana Twede. “Commercial amphoras: the earliest consumer packages?” In: *Journal of Macromarketing* 22.1 (2002), pp. 98–108.

[3] Jacques Yves Cousteau. “Fish men discover a 2,200-year-old Greek ship”. In: *National geographic* 105.1 (1954), pp. 1–36.

[4] Anthony J Parker. “Shipwrecks and ancient trade in the Mediterranean”. In: (1984).

[5] Mohamed Elgendy. *Deep Learning for Vision Systems*. Manning Publications, 2020.

[6] Waseem Rawat and Zenghui Wang. “Deep convolutional neural networks for image classification: A comprehensive review”. In: *Neural computation* 29.9 (2017), pp. 2352–2449.

[7] Zhong-Qiu Zhao et al. “Object detection with deep learning: A review”. In: *IEEE transactions on neural networks and learning systems* 30.11 (2019), pp. 3212–3232.

[8] Li Liu et al. “Deep learning for generic object detection: A survey”. In: *International journal of computer vision* 128.2 (2020), pp. 261–318.

[9] Yongcheng Jing et al. “Neural style transfer: A review”. In: *IEEE transactions on visualization and computer graphics* 26.11 (2019), pp. 3365–3385.

[10] Ian J Goodfellow et al. “Generative adversarial networks”. In: *arXiv preprint arXiv:1406.2661* (2014).

[11] Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. “Deep face recognition”. In: (2015).

[12] Ronald Poppe. “A survey on vision-based human action recognition”. In: *Image and vision computing* 28.6 (2010), pp. 976–990.

[13] Alexander Toshev and Christian Szegedy. “Deeppose: Human pose estimation via deep neural networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 1653–1660.

[14] Wei Niu, James Caverlee, and Haokai Lu. “Neural personalized ranking for image recommendation”. In: *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. 2018, pp. 423–431.

[15] L van der Maaten et al. “Computer vision and machine learning for archaeology”. In: (2007).

[16] Hongwei Qin et al. “When underwater imagery analysis meets deep learning: A solution at the age of big visual data”. In: *OCEANS 2015-MTS/IEEE Washington*. IEEE. 2015, pp. 1–5.

[17] Dario Lodi Rizzini et al. “Investigation of vision-based underwater object detection with multiple datasets”. In: *International Journal of Advanced Robotic Systems* 12.6 (2015), p. 77.

[18] Huimin Lu et al. “Underwater optical image processing: a comprehensive review”. In: *Mobile networks and applications* 22.6 (2017), pp. 1204–1211.

[19] Avi Abu and Roee Diamant. “A statistically-based method for the detection of underwater objects in sonar imagery”. In: *IEEE Sensors Journal* 19.16 (2019), pp. 6858–6871.

[20] Alan Gordon. “Use of laser scanning system on mobile underwater platforms”. In: *Proceedings of the 1992 Symposium on Autonomous Underwater Vehicle Technology*. IEEE. 1992, pp. 202–205.

[21] Md Moniruzzaman et al. “Deep learning on underwater marine object detection: A survey”. In: *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer. 2017, pp. 150–160.

[22] Pierre Drap et al. “Underwater photogrammetry and object modeling: a case study of Xlendi Wreck in Malta”. In: *Sensors* 15.12 (2015), pp. 30351–30384.

[23] Arthur L Samuel. “Some studies in machine learning using the game of checkers”. In: *IBM Journal of research and development* 3.3 (1959), pp. 210–229.

[24] Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O’Reilly Media, 2019.

[25] Andriy Burkov. *The hundred-page machine learning book*. Vol. 1. Andriy Burkov Canada, 2019.

[26] WJ Zhang et al. “On definition of deep learning”. In: *2018 World automation congress (WAC)*. IEEE. 2018, pp. 1–5.

[27] Ian Goodfellow et al. *Deep learning*. Vol. 1. 2. MIT press Cambridge, 2016.

[28] Warren S McCulloch and Walter Pitts. “A logical calculus of the ideas immanent in nervous activity”. In: *The bulletin of mathematical biophysics* 5.4 (1943), pp. 115–133.

[29] Frank Rosenblatt. *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory, 1957.

[30] Donald Olding Hebb. *The organization of behavior: A neuropsychological theory*. Psychology Press, 2005.

[31] Siegrid Lowel and Wolf Singer. “Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity”. In: *Science* 255.5041 (1992), pp. 209–212.

[32] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. *Learning internal representations by error propagation*. Tech. rep. California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

[33] Sebastian Ruder. “An overview of gradient descent optimization algorithms”. In: *arXiv preprint arXiv:1609.04747* (2016).

[34] Äaron Van Den Oord, Sander Dieleman, and Benjamin Schrauwen. “Deep content-based music recommendation”. In: *Neural Information Processing Systems Conference (NIPS 2013)*. Vol. 26. Neural Information Processing Systems Foundation (NIPS). 2013.

[35] Ronan Collobert and Jason Weston. “A unified architecture for natural language processing: Deep neural networks with multitask learning”. In: *Proceedings of the 25th international conference on Machine learning*. 2008, pp. 160–167.

[36] David H Hubel. “Single unit activity in striate cortex of unrestrained cats”. In: *The Journal of physiology* 147.2 (1959), pp. 226–238.

[37] David H Hubel and Torsten N Wiesel. “Receptive fields of single neurones in the cat’s striate cortex”. In: *The Journal of physiology* 148.3 (1959), pp. 574–591.

[38] David H Hubel and Torsten N Wiesel. “Receptive fields and functional architecture of monkey striate cortex”. In: *The Journal of physiology* 195.1 (1968), pp. 215–243.

[39] Yann LeCun et al. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

[40] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems* 25 (2012), pp. 1097–1105.

[41] Christian Szegedy et al. “Going deeper with convolutions”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.

[42] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (2014).

[43] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[44] François Chollet. “Xception: Deep learning with depthwise separable convolutions”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1251–1258.

[45] Jie Hu, Li Shen, and Gang Sun. “Squeeze-and-excitation networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 7132–7141.

[46] 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*. 2015, pp. 3431–3440.

[47] Athanasios Voulodimos et al. “Deep learning for computer vision: A brief review”. In: *Computational intelligence and neuroscience* 2018 (2018).

[48] Niall O’Mahony et al. “Deep learning vs. traditional computer vision”. In: *Science and Information Conference*. Springer. 2019, pp. 128–144.

[49] Ebrahim Karami, Mohamed Shehata, and Andrew Smith. “Image identification using SIFT algorithm: Performance analysis against different image deformations”. In: *arXiv preprint arXiv:1710.02728* (2017).

[50] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. “Surf: Speeded up robust features”. In: *European conference on computer vision*. Springer. 2006, pp. 404–417.

[51] Edward Rosten and Tom Drummond. “Machine learning for high-speed corner detection”. In: *European conference on computer vision*. Springer. 2006, pp. 430–443.

[52] Mark Everingham et al. “The pascal visual object classes (voc) challenge”. In: *International journal of computer vision* 88.2 (2010), pp. 303–338.

[53] Olga Russakovsky et al. “Imagenet large scale visual recognition challenge”. In: *International journal of computer vision* 115.3 (2015), pp. 211–252.

[54] Tsung-Yi Lin et al. “Microsoft coco: Common objects in context”. In: *European conference on computer vision*. Springer. 2014, pp. 740–755.