ISSN 2830-9820

VisualComputing MAGAZiNE

Vol. 1, Issue 2
2023

Road Object Detection and distance estimation using depth monocular model

In this issue

Classification of Suspected Pulmonary Nodules Based on AI Approaches

The ImageNet Dataset designed for use in visual object recognition

Visual Computing Magazine

The Preface

Welcome to the second issue of Visual Computing Magazine, where we are proud to showcase the exceptional talents of Masters and Bachelors students who are pushing the boundaries between visual computing and artificial intelligence. In this issue, we are delighted to present a remarkable collection of cutting-edge research and innovation that spans diverse disciplines from computer vision to medical imaging.

This issue begins with "Road object detection and distance estimation using a monocular depth model" presented by Baroud Yasmine et al., who proposed two approaches for the detection of small road objects and the estimation the distance between the detected objects and the vehicle.

Advancing Healthcare Diagnosis, "Classification of suspected lung nodules based on AI approaches" by Y. Aghiles Koulal et al. is an essential contribution. By employing advanced AI models, the authors aim to provide automated analysis of lung nodules, enabling early detection of lung pathologies and improving patient care.

Empathy meets innovation with "Object Recognition System on a Tactile Device for Visually Impaired" by A. Souayah et al. Through their work, they illuminate a path toward a more inclusive world by developing a tactile device that enables visually impaired individuals to perceive and interact with their surroundings.

In the realm of sports analysis, "Recognition and Analysis of Sports Actions in Real-Time Video" by M. H. Diab. et al. takes the center stage. Their real-time video analysis opens up exciting possibilities for sports enthusiasts, coaches, and analysts to gain deeper insights into athletic performance.

"Towards a New Data Representation: GANs for Medical Images Segmentation" by K. Aoucher et al. investigates an innovative approach of exploiting generative adversarial networks (GANs) for a few-step learning of lung tumor segmentation models..

Thanks to the power of deep learning, Y. Hanafi et al. in "Melanoma Identification Using Deep Learning" offers a reliable way to identify melanoma, furthering advances in early detection and prognosis.

Finally, we delve into the foundations of visual object recognition with "The ImageNet Dataset Designed for Use in Visual Object Recognition" by Slimane Larabi. As a testament to the importance of comprehensive datasets, this study sheds light on their crucial role in advancing the field of computer vision.

Chief Editor: *Prof. Slimane LARABI*

Visual Computing: A quarterly magazine. Volume 1, Issue 2, 2023
***ISSN:* 2830-9820**

Visual Computing Magazine

Road Object Detection and distance estimation using depth monocular model

Baroud Yasmine, Bourzam Saad Allah, LAICHE Nacera, USTHB

Abstract.
Accurate detection and recognition of road objects, especially small objects, are crucial for autonomous driving. In this article, we have designed two approaches to solve the problem of detecting small road objects and estimating the distance between the detected objects and the vehicle's camera. The first approach is based on object detection without segmentation, where we optimized the YOLOv5s model. The second approach involves object detection with segmentation using the Mask R-CNN model with Detectron2.

Furthermore, we developed two different models for estimating the absolute distance of the detected objects. The first model utilizes the object detection of optimized model along with information from a monocular depth map using MiDaS. The second model is a deep learning instance segmentation model that extracts specific information from the masks of the detected objects and utilizes relative distances generated by MiDaS to estimate the absolute distance (m).

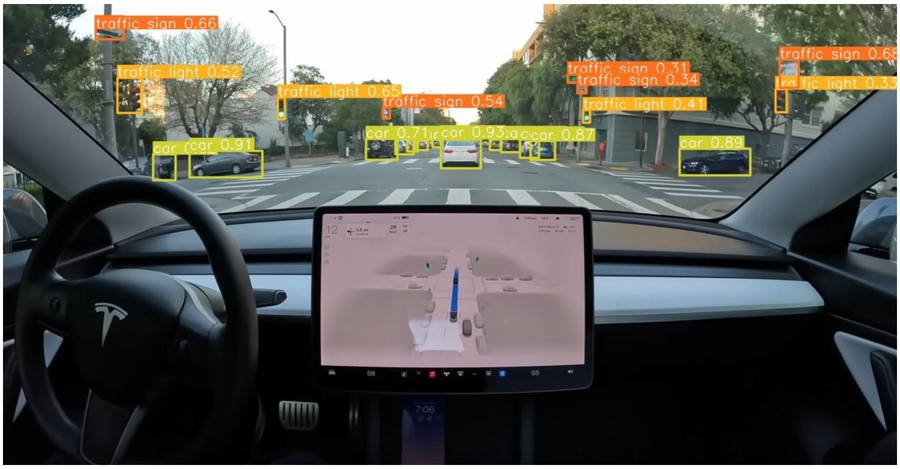


Figure 1. Example of result of road detection and recognition for a BDD10K-MV test image obtained with our optimized model.

1. Introduction

Real-time road object detection is critical for scene recognition and safe navigation of autonomous devices in natural environments. Furthermore, precise distance estimation is essential for advanced autonomous driving systems to provide safety features such as adaptive cruise control and collision avoidance. While radars and lidars can provide distance information, they are either costly or less accurate than image sensors when it comes to object information. However, detecting small objects and objects located at large distances in road images remains challenging [1], particularly using light models.

Additionally, distance estimation between the detected objects and vehicle’s camera is very challenging using 2D monocular camera. In this article, we have proposed two methods to solve the problem of detecting small road objects and estimating the distance between the detected objects and the vehicle's camera using a single image.

Road Object Detection and distance estimation using depth monocular model

Baroud Yasmine, Bourzam Saad Allah, LAICHE Nacera, USTHB

2. Method

We proposed two methods for designing a system that detects and recognizes road objects and estimates the distances between the detected objects and the camera using a single image.
Our design consists of three main parts:

2.1. Object Detection and Recognition:

- Method 1: We utilized our proposed model, based on a combination of YOLOv5s [2] and Swin transformer [3], for object detection without segmentation.
- Method 2: We employed the Mask R-CNN model [4] with Detectron2 [5] for segmentation-based object detection. The goal is to compare performance and determine the optimal execution speed for our real-time distance estimation system.

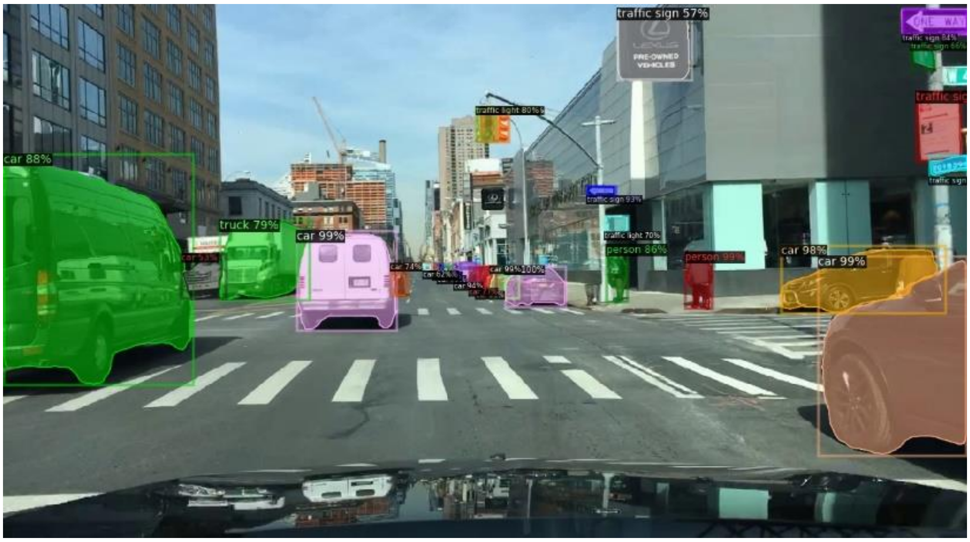


Figure 2. Example of result of road detection, segmentation and recognition for a BDD10K-MV test image obtained with our model trained Mask R-CNN with Detectron2..

2.2. Distance Estimation:

- We developed two different models for estimating the absolute distance of detected objects. The first model combines the object detection results with information from a monocular depth map generated by the MiDaS model. The second model utilizes deep learning instance segmentation (Mask R-CNN with Detectron2), specific mask information, and information from MiDaS [6].

Road Object Detection and distance estimation using depth monocular model

Baroud Yasmine, Bourzam Saad Allah, LAICHE Nacera, USTHB

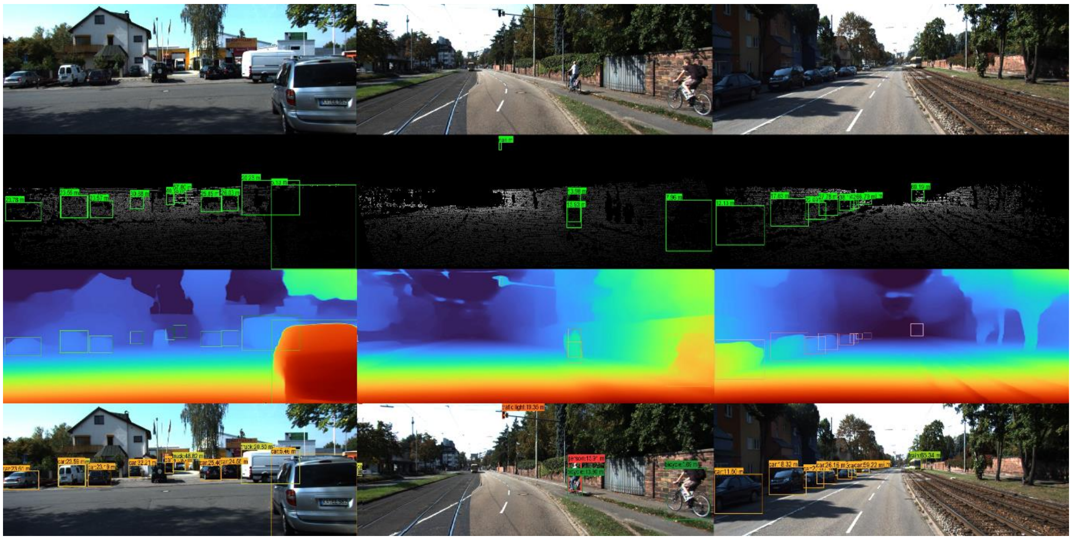


Figure 3. Examples of estimating distance using Mask R-CNN with Deetectron2 and MiDaS model on KITTI dataset.

The process involves running the two models in parallel. The MiDaS model predicts the relative depth map of the input image, while the proposed model locates and classifies road objects within the image. The location results of each object defined by bounding boxes obtained are overlaid on the estimated relative depth image. The relevant (relative) distance of an object is calculated by averaging the relative distance of all pixels not null within the defined bounding box in the MiDaS estimated depth map. To convert the relative distance (REV) of an object into real distance (ABS), the real distance of the objects in the images is estimated using LIDAR point clouds provided as ground truths in the KITTI Raw dataset. Then, the relationship between the real (absolute) distance and the relative distance is calculated using a quadratic mathematical formula that minimizes the error between the actual distance and the relative distance.

For the second approach mentioned above, we followed the same steps, but instead of considering the points inside the bounding box generated by the proposed model, we considered the points inside the object mask detected and segmented by the Mask R-CNN model with Detectron2.

2.3. Data augmentation:

To address the limited availability of data for certain road object classes, we proposed two methods for data augmentation. The first method involves applying data augmentation techniques such as photometric distortion, geometric distortion, etc, making the models more robust

Road Object Detection and distance estimation using depth monocular model

Baroud Yasmine, Bourzam Saad Allah, LAICHE Nacera, USTHB

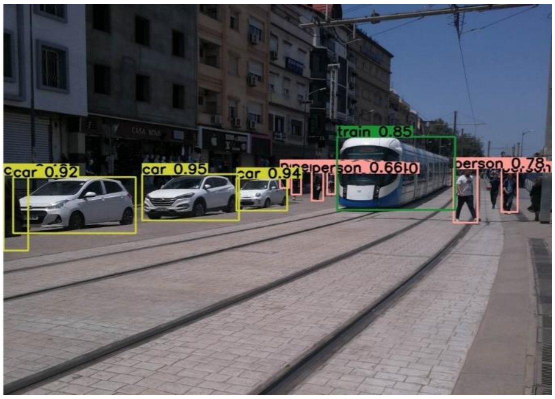


Figure 4 Example of Road object detection and recognition result for an image of Algiers obtained by our model optimized.

against images from different environments and lighting conditions. The second method aims to increase the number of instances for classes with low representation in the BDD10K dataset, strengthening the object representation and achieving better balance among classes. We refer to this new dataset as BDD10K-MV (BDD10K-Mapillary Vistas) in our article.

3. Experimental results

In this section, we present the results of our experiments conducted to evaluate the performance of the proposed approaches. Firstly, we evaluated the proposed model through various experiments using the KITTI dataset toward six different road objects (person, bicycle, car, motorcycle, bus, train, truck) and our newly created dataset BDD10K-MV with nine road object (Person, Traffic Sign, Traffic light, Car, Truck, Motorcycle, Train, Bus Bicycle). The evaluation metrics focused on mean average precision (Map50%) and revealed that the proposed model achieved an impressive performance with a Map50% score of 95% on the KITTI dataset and 56.4% on the BDD10K-MV dataset. Compared to YOLOv5s, our approach improves mAP by 1.5% and 3.5% on the KITTI and BDD10K-MV datasets respectively, enabling higher accuracy in detecting small objects road images and objects located at a large distance.

Comparatively, we assessed the Mask R-CNN model with Detectron2, which achieved an average precision (AP50%) of 56.88% on the BDD10K-MV dataset.

The estimation of absolute distance was evaluated using all images from the KITTI Raw dataset, and the proposed models demonstrated their effectiveness, which makes the solution highly competitive with existing approaches. The choice of model ultimately depends on specific needs and objectives. For our focus on precise distance estimation and real-time detection in autonomous vehicles, we selected the proposed model of detection and recognition road objects and MiDaS variant of dpt_swin2_l384 as the preferred detection and recognition with distance estimation model.

Road Object Detection and distance estimation using depth monocular model

Baroud Yasmine, Bourzam Saad Allah, LAICHE Nacera, USTHB

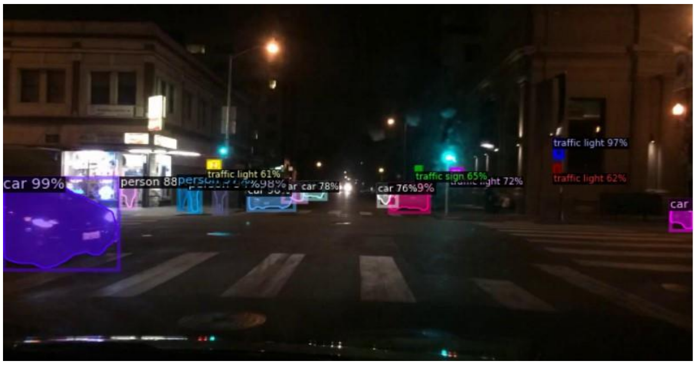


Figure 5. Detection, recognition and segmentation results for road objects in a BDD10K test image obtained by the Mask R-CNN model with Detectron2

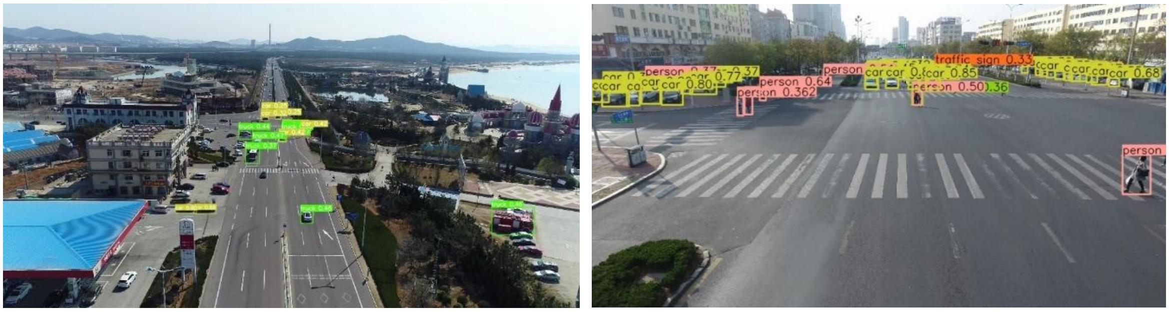


Figure 6. Road object detection and recognition results on two images from the UAVDT UAVDT_benchmark drone dataset obtained with our optimized model, showing the effectiveness of our model in detecting small objects.

References

[1] Armin Masoumian et al. “Absolute Distance Prediction Based on Deep Learning Object Detection and Monocular Depth Estimation Models”. In: (oct. 2021). url : https://arxiv.org/abs/2111.01715

[2] Ultralytics. “You Only Look Once : Unified, Real-Time Object Detection”. In : (2022). url : https://github.com/ultralytics/yolov5/issues/6998.

[3] Ze Liu et al. “Swin Transformer : Hierarchical Vision Transformer using Shifted Windows”. In : (2021). arXiv : 2103.14030 [cs.CV].

[4] Kaiming He et al. “Mask R-CNN”. In : (2018). arXiv : 1703.06870 [cs.CV].

[5] Meta AI. Detectron2 : A PyTorch-based modular object detection library. consulté en avril 2023. url : https://youtu.be/egsoXN-xjAo

[6] René Ranftl et al. “Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer”. In : (2020). arXiv : 1907.01341 [cs.CV].

Visual Computing Magazine

Classification of Suspected Pulmonary Nodules Based on AI Approaches

Yidhir Aghiles Koulal, Amayas Labchri, Djamila Dahmani, USTHB

Abstract

Our research is focused on utilizing artificial intelligence techniques to classify pulmonary nodules, with the goal of improving diagnostic accuracy and efficiency. By employing sophisticated models, we aim to provide automated analysis of pulmonary nodules, enabling early detection of lung pathologies and enhancing patient care.

To achieve this, we have developed a ground breaking approach that starts with pre-segmented 3D CT scans of nodules. From these scans, we extract three meticulously crafted 2D images representing the nodule from different perspectives: x, y, and z views. Taking inspiration from the pioneering work of G. Hinton [1] which was further improved by Sara Sabur et al in 2017 [2], our methodology revolves around the implementation of capsule networks. As a result, we have created three specialized models, each tailored to a specific view.

In cases where there is disagreement among the three models, we employ three basic methods for final classification: strict classification, majority classification, and threshold-based classification. Furthermore, we have incorporated game theory principles to determine the category of the nodule (benign or malignant).

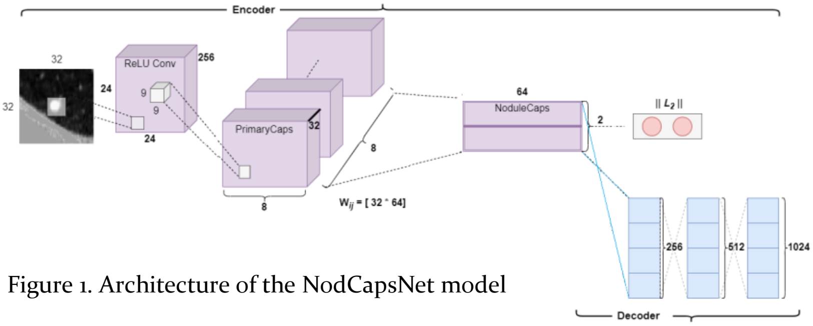


Figure 1. Architecture of the NodCapsNet model

1. Introduction

Our research uses AI techniques to classify pulmonary nodules for improved diagnostic accuracy and efficiency. Our groundbreaking approach starts with pre-segmented 3D CT scans, from which we extract three 2D images representing the nodule from different perspectives. Our methodology revolves around the implementation of capsule networks, resulting in three specialized models tailored to specific views. We use three basic methods for final classification and incorporate game theory principles to determine the nodule's category.

2. Method

2.1. NodCapsNet Model

The dataset used to train and test our neural network is the "Data Science Bowl 2017" dataset, abbreviated as DSB2017 [3]. The dataset consists of approximately 753 segmented nodules of size (64*64*64).

a) Encoding:

The encoder plays a crucial role in transforming a 32x32 pixel image of a nodule into a 64-dimensional instantiation parameter vector that encapsulates essential nodule information.

Classification of Suspected Pulmonary Nodules Based on AI Approaches

Yidhir Aghiles Koulal, Amayas Labchri, Djamila Dahmani, USTHB

The convolutional layer is responsible for detecting basic features in the input image using 256 kernels of size 9x9. The primary capsules, eight in total, each apply 32 kernels of size 9x9x256 to detect specific nodule features such as size, texture, and orientation.

The NoduleCaps layer consists of two capsules of size 64. This layer encodes not only the activation of features but also the spatial relationships between these features.

The capsules then generate output vectors specific to different nodule classes, such as "benign" or "malignant," by utilizing vector output norms and the squash function.

b) Decoding:

After encoding, the decoding process is necessary to encourage the capsules to select the most relevant features.

We apply deconvolution layers to progressively enlarge the vector representation and restore the spatial features of the image.

Our goal is to generate a reconstructed image that closely resembles the original image, with the Euclidean distance serving as the loss function. The closer the reconstructed image is to the input image, the better the decoding process.

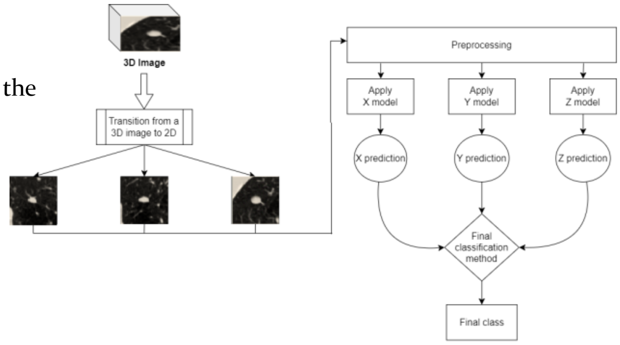


Figure 2. General diagram of the classification system

c) Integration of Gabor Filter:

The Capsule Network model for pulmonary nodule classification has been enhanced through the integration of the Gabor filter [4]. In standard capsule networks, visual features are learned automatically from the input data, limiting the ability to explicitly select features like size and orientation. By incorporating the Gabor filter, originally developed for tasks such as edge detection and texture analysis, the network gains the ability to capture more relevant visual attributes from the nodule images. The Gabor filter operates by convolving the input image with specific kernels designed to extract spatial frequency, orientation, and phase information. This integration allows the capsule network to capture subtle patterns and variations in the nodules, leading to improved classification accuracy and performance.

Classification of Suspected Pulmonary Nodules Based on AI Approaches

Yidhir Aghiles Koulal, Amayas Labchri, Djamila Dahmani, USTHB

2.2 Classification system

Phase 1:

The system takes a 3D image (646464) of a segmented pulmonary nodule from a lung CT scan as input. To optimize the classification process, we select the central slice for each view axis. This reduces complexity and speeds up the training process.

Phase 2:

In the second step of our system, we apply preprocessing techniques to our 2D images, such as intensity normalization, resizing, and Gaussian filtering to reduce noise. These processed images are then fed into our pre-trained model to generate predictions.

Phase 3:

The final phase of the system involves making a conclusive decision about the nature of the input nodule based on the three predictions obtained in Phase 2. To achieve this, we employ a final classification method.

Final Classification Methods

Majority Classification:

The class that is predicted most frequently by the three models is selected as the system's final decision.

Strict Classification:

If at least one of the models predicts a malignant class, that class is returned as the final decision.

Threshold-based Classification with Coefficients:

The predictions of the three models are weighted with coefficients, where the model with the highest accuracy has the highest coefficient. Their average is then compared to a threshold calculated during the training process.

Game Theory Classification:

The system creates a zero-sum game where it confronts the views as players, and their strategies are the features calculated from a set of reference images.

In our game theory model, we have utilized the following features: energy and homogeneity from the co-occurrence matrix, as well as Hausdorff distance and spectral radius.

3. Experimental Results

Excluding the Gabor filter led to a decrease in loss but a low precision rate. However, incorporating this filter significantly improved the precision rate of the model across all three viewing angles, achieving a minimum of 92%. Furthermore, enhancing the number and size of primary capsule vectors yielded better outcomes with fewer iterations, thereby reducing the duration of training (see table 1).

Based on these findings, we have chosen NodCapsNet32 as the deep learning model for our application.

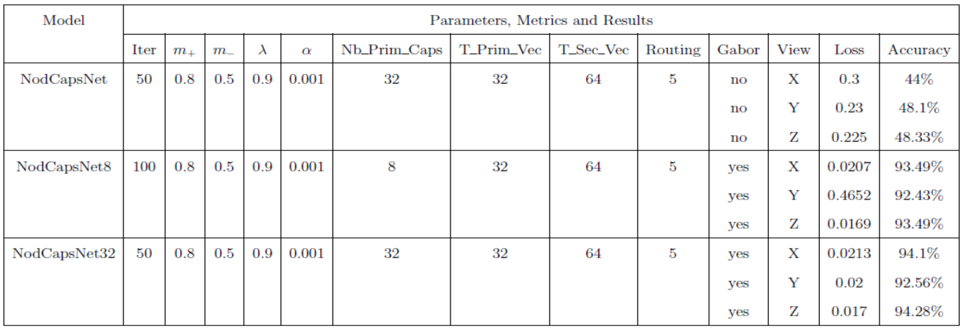
Visual Computing Magazine

Classification of Suspected Pulmonary Nodules Based on AI Approaches

Yidhir Aghiles Koulal, Amayas Labchri, Djamila Dahmani, USTHB

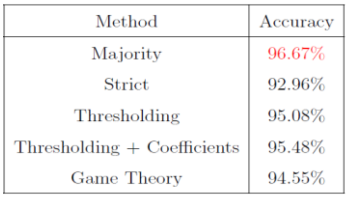
The majority voting method had the highest precision rate, as all three models had high accuracy for the viewing angles they were trained on. The thresholding model also produced good results due to the calculation of credible thresholds during training. This approach improved the overall performance of the model by using reliable thresholds for calculating prediction averages.

Using game theory, we achieved satisfactory results by utilizing features of lung nodules from a set of pre-selected reference images, with a prediction rate of 85%. However, it is important to note that the similarity between nodules can affect the results, as a benign nodule may have similar characteristics to a malignant one. For the strict method, forcing the models to agree did not help with decision-making, as a model trained on one view may interpret the nodule class differently.



Model	Parameters, Metrics and Results											
	Iter	m_+	m_-	λ	α	Nb_Prim_Caps	T_Prim_Vec	T_Sec_Vec	Routing	Gabor	View	Loss	Accuracy
NodCapsNet	50	0.8	0.5	0.9	0.001	32	32	64	5	no	X	0.3	44%
										no	Y	0.23	48.1%
										no	Z	0.225	48.33%
NodCapsNet8	100	0.8	0.5	0.9	0.001	8	32	64	5	yes	X	0.0207	93.49%
										yes	Y	0.4652	92.43%
										yes	Z	0.0169	93.49%
NodCapsNet32	50	0.8	0.5	0.9	0.001	32	32	64	5	yes	X	0.0213	94.1%
										yes	Y	0.02	92.56%
										yes	Z	0.017	94.28%

Table 1. NodCapsNet model results



Method	Accuracy
Majority	96.67%
Strict	92.96%
Thresholding	95.08%
Thresholding + Coefficients	95.48%
Game Theory	94.55%

Table 2. Results of classification methods

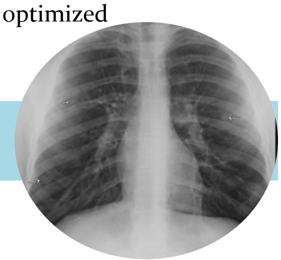
References:

[1] G. E. Hinton, A. Krizhevsky S. D. Wang , Transforming Auto-encoders, University of Toronto. http ://www.cs.toronto.edu/ fritz/absps/transauto6.pdf

[2] Sara Sabour et ul, Dynamic Routing Between Capsules. 2017. https ://doi.org/10.48550/arXiv.1710.09829.

[3] Kaggle. Data Science Bowl 2017. https ://www.kaggle.com/c/data-science-bowl-2017

[4] J. G. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. 1985. https ://doi.org/10.1364/JOSAA.2.001160



Object Recognition System on a Tactile Device for Visually Impaired

Souayah Abdelkader, MOKRETAR KRAROUBI Abderrahmene, Slimane LARABI, USTHB

Abstract

People with visual impairments face numerous challenges when interacting with their environment. Our objective is to develop a device that facilitates communication between individuals with visual impairments and their surroundings. The device will convert visual information into auditory feedback, enabling users to understand their environment in a way that suits their sensory needs.
Initially, an object detection model is selected from existing machine learning models based on its accuracy and cost considerations, including time and power consumption. The chosen model is then implemented on a Raspberry Pi, which is connected to a specifically designed tactile device. When the device is touched at a specific position, it provides an audio signal that communicates the identification of the object present in the scene at that corresponding position to the visually impaired individual.
Conducted tests have demonstrated the effectiveness of this device in scene understanding, encompassing static or dynamic objects, as well as screen contents such as TVs, computers, and mobile phones.

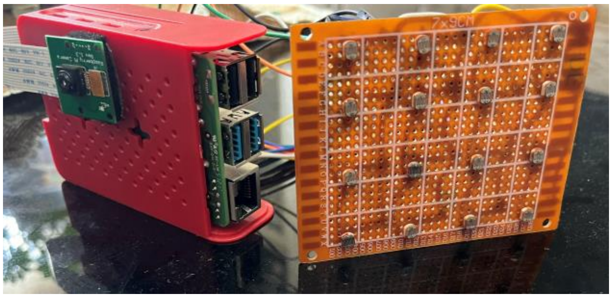


Figure 1. The developed prototype

1- Introduction

People with visual impairments face numerous challenges in their daily lives. They are unable to perceive the world in the same way as those with sight and encounter multiple difficulties, including orientation, obstacle detection and avoidance, limited mobility, and an inability to recognize shapes and colors of objects in their surroundings. In addition to these challenges, they are completely excluded from understanding and interacting with the real world scene.
Numerous technological advancements have been made to assist people with visual impairments. Among the different technological solutions deployed to address this specific need, computer vision-based solutions appear as one of the most promising options due to their affordability and accessibility.
Systems with human-scene interaction generate outputs after processing the captured scene. They consist of a set of computer vision and machine learning techniques aimed at improving the user's life in various activities such as content interpretation, navigation, etc. Generally, these systems process the data received from the real world using depth or RGB sensors and transform them into instructions and signals [1,2].
The goal of this work is to assist individuals with visual impairments in perceiving the information contained in an image by displaying the coded scene on a tactile device. They can explore the image by touching the pins on the device, with each pin representing a corresponding object in the scene.
The developed prototype is illustrated in Figure 1.

Object Recognition System on a Tactile Device for Visually Impaired

Souayah Abdelkader, MOKRETAR KRAROUBI Abderrahmene, Slimane LARABI, USTHB

2- Method

Our system aims to assist visually impaired individuals in identifying objects and their locations from images. A tactile device has been developed to provide auditory feedback corresponding to the identity of the detected object, thereby helping these individuals obtain information about the scene (see Figure 1). The proposed system is capable of identifying 17 types of objects in the observed scene.

This system is divided into three processes as illustrated in Figure 2.

The first two processes cooperate in interpreting and detecting objects in the observed scene. Since the system is embedded on a Raspberry Pi, it has limited resources (low RAM and processing power). Therefore, we have developed three object detection models, each responsible for detecting objects in a specific environment: Office, Kitchen, and Bedroom.

The main process is responsible for recognizing the appropriate environment in order to load the corresponding model. This process is reactivated at each new camera location and when the detection rate falls below a predefined threshold.

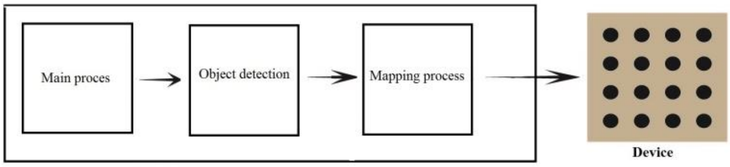


Figure 2. The processes of the proposed system

The second process is responsible for detecting objects and their locations in the image, and it transfers the coordinates of the object locations to the next process.

The third process involves associating the detected objects with a location on the tactile device and interacting with the user to produce corresponding sound feedback for the detected object.

2.1-The Detection Processes

The acquired image is input into the main process, which determines the appropriate environment or scene category (e.g., office, kitchen, bedroom) based on the visual cues and characteristics present in the image.

These three models are based on YOLOv5 and have been retrained on a dataset consisting of seven specific object classes. The goal of each model is to detect and recognize objects belonging to these seven classes in their corresponding environment.

The system operates using an object detection model that is responsible for detecting characteristic objects in each environment. Then, the k-nearest neighbors algorithm is executed to recognize the observed environment. In this algorithm, objects represent the features, and environments represent the target classes. Once the appropriate environment is determined, the second process takes over for object detection and location, while the first process is paused. It transfers the coordinates of each detected object to the final process. This process also has the responsibility of reactivating the main process when the detection rate falls below a predefined threshold (e.g., when the object detection model fails to detect more than 20% of the objects in the observed environment).

Object Recognition System on a Tactile Device for Visually Impaired

Souayah Abdelkader, MOKRETAR KRAROUBI Abderrahmene, Slimane LARABI, USTHB

2.2-Mapping process

This process is responsible for converting the coordinates of objects in the image into relative coordinates on the tactile device. In cases where there is overlap between two objects, where both objects may appear in the same grid cell of the tactile device, or when a relatively large object occupies multiple grid cells, we have developed an algorithm to determine the order of objects belonging to the same grid cell. Additionally, this process is responsible for interacting with the user through the tactile device. It produces sound feedback corresponding to each detected object. When the user touches or interacts with a specific pin on the tactile device, a specific sound is emitted to provide feedback to the user.

2.3-Model selection

In order to select the appropriate model for the specific task of integrating an object detection model into a Raspberry Pi, we conducted a comparative study of object detection algorithms based on Convolutional Neural Networks (CNNs). Considering our objective of achieving acceptable precision and recall values while working with embedded systems, we conducted the comparison while considering the following constraints:

We focused on using the latest reduced versions of each model, commonly referred to as "tiny models," such as YOLOv5 [4], Faster R-CNN [5], and SSD [6].

The object detection models used in the comparison were trained on the same image dataset and shared the same backbone architecture. This ensured that we could make meaningful observations regarding the advantages and disadvantages of these methods.

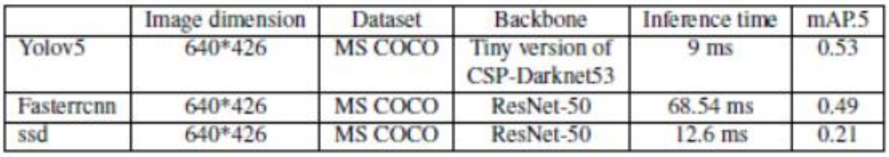


	Image dimension	Dataset	Backbone	Inference time	mAP.5
Yolov5	640*426	MS COCO	Tiny version of CSP-Darknet53	9 ms	0.53
Fasterrcnn	640*426	MS COCO	ResNet-50	68.54 ms	0.49
ssd	640*426	MS COCO	ResNet-50	12.6 ms	0.21

Table 1. Comparison results on public database

Tables 1 and 2 present the results obtained on the MS-COCO benchmark and a collected image dataset in terms of mAP0.5 (mean Average Precision at IoU threshold of 0.5). These results validate the effectiveness of YOLOv5 in comparison to Faster R-CNN and SSD. The tables demonstrate that the YOLOv5 structure is better suited for real-time applications due to its faster processing speed compared to the other structures.

The selected model, determined by the main process, utilizes a YOLOv5-based object detection method to identify and locate objects in the image. It generates bounding boxes that enclose each detected object, accompanied by confidence scores that must exceed 0.5 to be deemed valid.

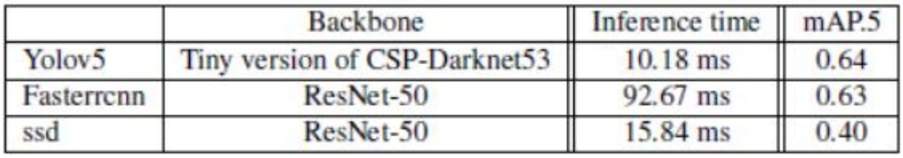
The coordinates of the detected objects, represented by the bounding boxes, are extracted from the object detection model and transmitted to the final process for encoding them on the tactile device.

Object Recognition System on a Tactile Device for Visually Impaired

Souayah Abdelkader, MOKRETAR KRAROUBI Abderrahmene, Slimane LARABI, USTHB

3- Experimental Results

In Figure 3, we present the components of our interactive device designed for visually impaired individuals.
This compact and portable device is a Raspberry Pi equipped with a high-definition camera, a 2GB CPU, and RAM. The device analyses the user's environment and detects objects in real-time. The gathered information is then transmitted to the user through a haptic feedback system.
The haptic feedback system utilizes a device with 16 photoresistor sensors, enabling visually impaired individuals to comprehend their environment using their fingers. These sensors detect the presence of fingers and convert this information into audio feedback.



	Backbone	Inference time	mAP.5
Yolov5	Tiny version of CSP-Darknet53	10.18 ms	0.64
Fasterrcnn	ResNet-50	92.67 ms	0.63
ssd	ResNet-50	15.84 ms	0.40

Table 2. Comparison results on our collected images.

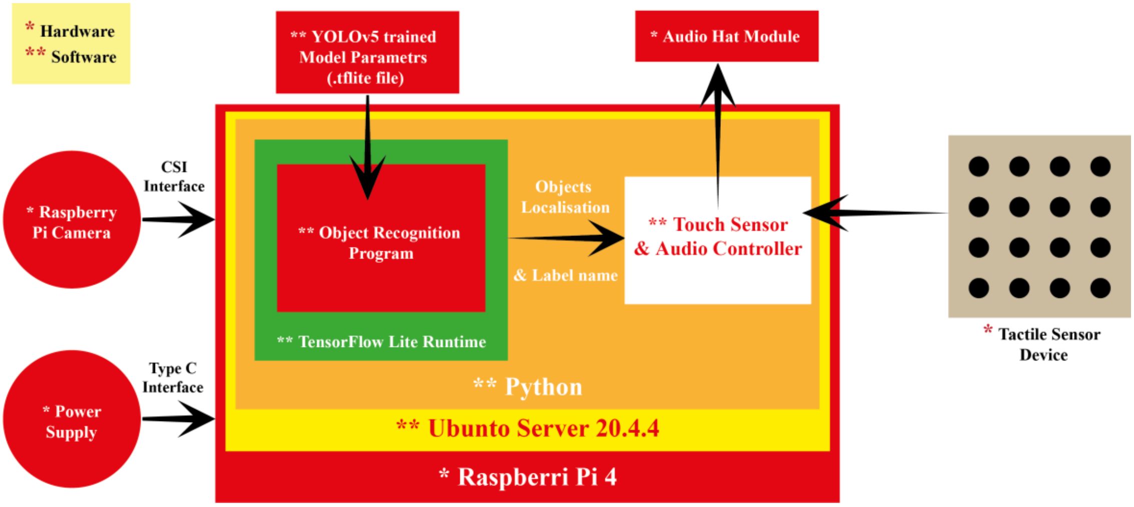


Figure 3. Components of the system

Object Recognition System on a Tactile Device for Visually Impaired

Souayah Abdelkader, MOKRETAR KRAROUBI Abderrahmene, Slimane LARABI, USTHB

3.1 Making the device

Our main goal is to enable tactile-audio interaction with visually impaired users. To accomplish this, we have opted to utilize photoresistor technology, an electronic component that exhibits varying electrical resistance in response to incident light. The resistance of a photoresistor changes inversely proportional to the intensity of light it receives. In tactile interaction, this technology can be employed to detect changes in light caused by the user's touch on a light-sensitive surface.

By arranging multiple photoresistors as pins on a surface, we can detect which resistors are touched by the user. This enables tactile interaction where the user can interact with different pins and trigger actions, such as audio feedback through an audio output module connected to the Raspberry Pi. Each touched resistor corresponds to a specific sound based on the detected object's position in the image relative to the pin.

Figure 4 illustrates the circuit connecting 16 photoresistors, with each photoresistor connected to one of the Raspberry Pi's pins. The associated code utilizes the RPi.GPIO library to manage GPIO pins on the Raspberry Pi. It configures the port for the photoresistor as an input. In the main loop, it checks the state of the photoresistor. If it is triggered (HIGH), it displays a message indicating that the user has touched the photoresistor.

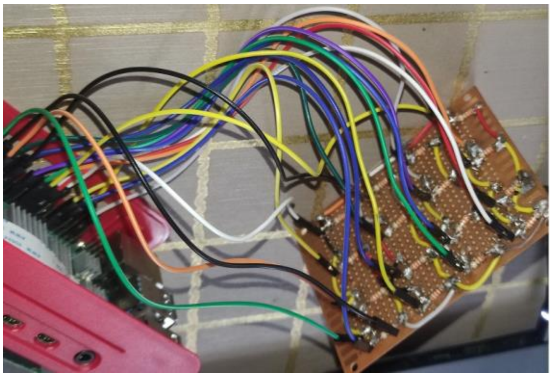


Figure 4. Connecting Raspberry Pi and the tactile device.

3.2 Object Detection

The primary objective of this project is to identify and detect 17 different classes distributed across three categories: office, kitchen, and bedroom. During the data collection phase, we obtained a total of 2677 images specifically for the office environment.

The dataset consists of a total of 2677 samples, which are divided into 7 classes representing office environments. The smallest class contains approximately 290 samples. Each class has an adequate number of images distributed across the training, validation, and test sets. The image dataset is organized into three files: train (70%), validation (20%), and test (10%).

3.3 Transfer Learning

We fine-tuned and configured the YOLOv5 architecture specifically for our dataset. To achieve this, we employed transfer learning, adapting the YOLOv5 framework to be compatible with our dataset. We utilized pre-trained weights from a different model that had been trained on the extensive COCO dataset.

For training our model (yolov5s.pt), we utilized the standard Colab VM with 12GB of GPU memory. To enhance the robustness of the trained model and better utilize the available GPU resources, we set the batch size to 4. Additionally, we conducted training for a total of 100 epochs, observing that the trained model reached stability.

Throughout the experiments, we incorporated various hyperparameters. Some of these included weight decay = 0.0005, initial learning rate = 0.0042, final learning rate = 0.1, and momentum = 0.937. These parameters were maintained at their default values. Ultimately, we trained and tested YOLOv5 on the Colab VM using our dataset.

3.4 Results of objects detection

To provide a more detailed analysis of the model's training process and performance, Figure 5 (left) displays a plot showcasing the precision and recall mapping for detecting the seven classes during training. From the figure, it is evident that the model achieved a mean Average Precision (mAP) of 86.3%. This mAP value represents the area under the curve, indicating the trained model's ability to accurately detect objects with high precision and recall values.

To highlight the superiority of our selected object detection model for the desk environment (comprising the previously mentioned 7 classes), we conducted a comparison with the detection model prior to transfer learning, namely yolov5s. Figure 5 showcases the mAP results obtained by both models on a training image set.

It is evident from the results that our model surpasses yolov5s in terms of average precision for the 7 classes. It is important to highlight that the object labeled as dining table in yolov5 is distinct from the desk object. Based on the conducted experiments, our transfer learning model derived from yolov5s exhibits superior performance. Therefore, we can confidently utilize our model for the project.

3.5 Mapping

The pin grid provides an organized structure and spatial reference for each object based on its position and size. This facilitates further processing or interaction with the detected objects within the project's context.

The detected objects in the image are associated to their corresponding cells. Each object's bounding box is associated to the grid cell as long as the majority of its surface is within that cell. A bounding box can be associated with multiple cells, and likewise, a cell can have multiple bounding boxes .

Object Recognition System on a Tactile Device for Visually Impaired

Souayah Abdelkader, MOKRETAR KRAROUBI Abderrahmene, Slimane LARABI, USTHB

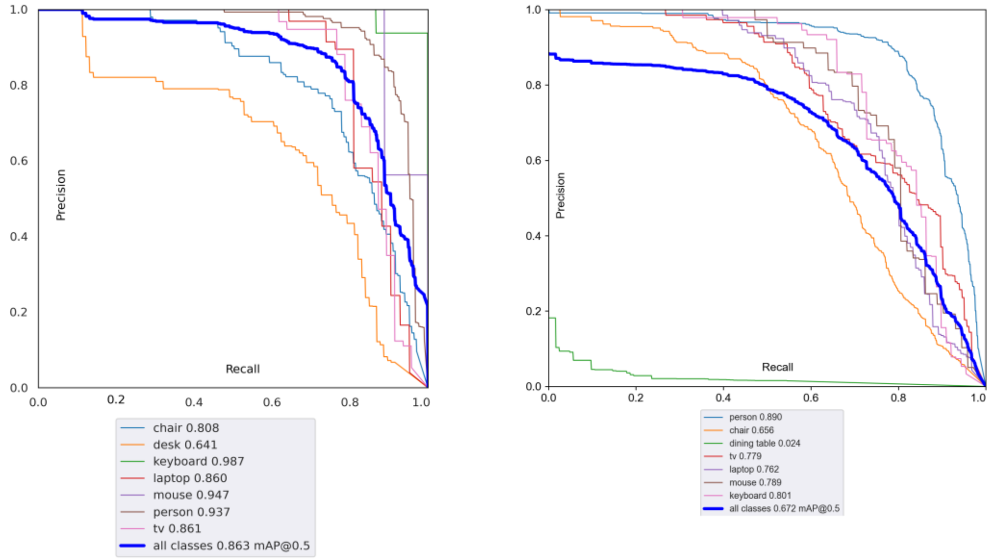


Figure 5. Precision, Recall for our model (left) and yolov5s (right) on train data.

References

[1] Zatout, Chayma and Larabi, Slimane and Mendili, Ilyes and Barnabé, Soedji Ablam Edoh.Ego-Semantic Labeling of Scene from Depth Image for Visually Impaired and Blind People. IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019, pp. 4376-4384.
[2] Larabi S. Zatout C. Semantic scene synthesis: application to assistive systems. The Visual Computer, 38:2691-2705, 2022.
[3] Tsung-Yi et al. Lin. Microsoft coco: Common objects in context. In Computer Vision – ECCV 2014, pages 740–755, Cham, 2014. Springer International Publishing
[4] Chien-Yao Wang, Alexey Bochkovskiy, and Hong Yuan Mark Liao. Scaled-yolov4: Scaling cross stage partial network, 2021.
[5] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, Advances in Neural Information Processing Systems, volume 28. Curran Associates,
Inc., 2015.
[6] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: Single shot MultiBox detector. In Computer Vision – ECCV 2016, pages 21–37. Springer International Publishing, 2016.

Recognition and analysis of sports actions in real-time video

DIAB M. Hicham, BOUTICHE Ahmed Imed, LAICHE Nacera, USTHB

Abstract
In this study, a deep learning model is developed for real-time and pre-recorded classification of 16 distinct fitness exercises. The model leverages the Mediapipe Pose estimation model to extract key body landmarks, which serve as training data for accurate exercise pose prediction. The extracted landmarks also enable a comprehensive analysis of the exercises, including counting exercise repetitions, providing guidance on the next move, and offering posture tips. The proposed approach demonstrates promising results in effectively classifying fitness exercises while enhancing monitoring and guidance capabilities.

1. Introduction

Regular exercise is essential for maintaining overall health and well-being. As the popularity of fitness training increases, there is a need for automated systems that can accurately classify and analyse different exercises. In this study, we develop a deep learning model to classify 16 fitness exercises. By utilizing the Mediapipe Pose estimation model, key body landmarks are extracted for training the model. Our goal is to provide real-time feedback, monitor progress, and offer personalized guidance to individuals during their fitness routines. Additionally, we perform a detailed analysis of the exercises using the extracted landmarks. This research aims to enhance monitoring and guidance in fitness training programs, promoting improved outcomes and overall well-being.

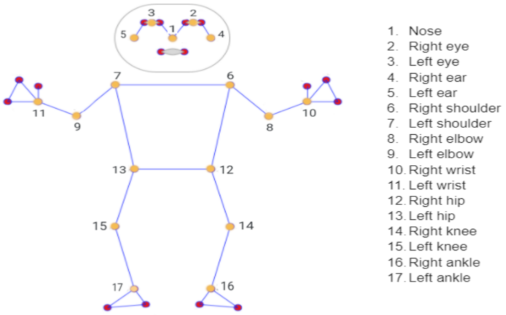


Figure 1: List of key points in our approach

2. The method

In this section, we will discuss the step-by-step process of our detailed approach. First, we will cover the extraction of 2D pose data. We will then delve into the normalization of this data and the subsequent building and training of a neural network using the extracted pose data. In the second part, we will shift our focus to the analysis of exercises based on the 2D pose data.

2.1 Exercise Recognition

Feature Extraction

In the feature extraction stage, we chose to utilize the Mediapipe Pose model [1] due to its user-friendly nature and excellent performance. The Mediapipe Pose model accurately estimates key body landmarks, such as shoulders, elbows, wrists, hips, knees, and ankles. Its ease of use and reliable performance make it an ideal choice for extracting 2D pose data from the collected dataset of videos containing 16 workout exercises.

Recognition and analysis of sports actions in real-time video

DIAB M. Hicham, BOUTICHE Ahmed Imed, LAICHE Nacera, USTHB

During the analysis of our collected videos, we observed that the key points of inner eyes and mouth did not provide relevant information for our objective of analysing and recognizing human actions. Their presence in our data did not add significant value to our data extraction process or our subsequent analysis steps. Hence, we excluded these key points from our feature extraction process, focusing solely on the essential body landmarks for exercise classification and analysis.

We extracted 2D pose data from videos of 16 exercises: 10 exercises generated by infinity.ai [2] as 3D synthetic video data and the remaining 6 exercises from YouTube videos. The distribution of the extracted 2D pose data is given by figure 2.

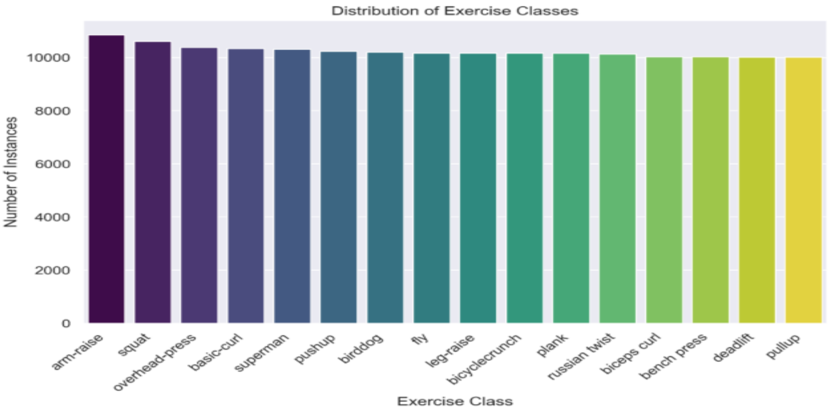


Figure 2. Distribution of extracted 2D pose data

Data Normalization

Data normalization [3] is crucial for ensuring the comparability and consistency of 2D pose data. It eliminates biases related to size and position, allowing accurate analysis and interpretation, regardless of pose or individual placement within the frame. Normalizing coordinates based on the center of gravity and body length is key to achieving this.

Model Architecture

Our neural network is designed to classify 16 different workout exercises based on 2D Mediapipe Pose data. The architecture of our model consists of several key components that enable accurate exercise recognition. The model architecture is given by figure 3.

Model Usage Method

One difficulty when using neural network models trained on Mediapipe Pose data is the rapid and fluctuating predictions. This occurs because the models are trained on static images, while the prediction task involves dynamic videos with rapidly changing poses. [4]

To address the issue, we use a queue-based smoothing technique. We store the predictions for each frame in a queue and select the most frequent prediction. This effectively smooths the prediction process and reduces rapid changes.

2.2 Exercise Analyse

Exercise analysis during execution is crucial for optimizing athlete performance by providing technical guidance and evaluating progress. The analysis includes:
-Repetition Calculation: Measures endurance progression and sets specific goals for the athlete's improvement.
-Issuing Instructions: Offers clear instructions indicating the next movement to perform during the exercise.
-Providing Relevant Tips: It provides guidance on making minor adjustments to your position, indicating if your posture is incorrect during certain exercises.
-Timer Function: Controls exercise duration for performance evaluation.

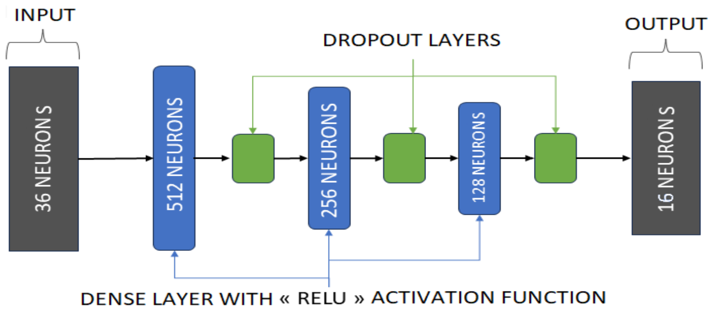


Figure 3. Model architecture

Example of Queue-Based Smoothing

For instance, in a scenario with “plank” and “push-up” pose classification, using a queue length of 5, we consider the five previous predictions. The most frequent prediction in the queue is chosen as the final prediction. Figure 4 illustrates an example. By implementing this approach with a queue length of 25, corresponding to the average frame frequency of most videos and webcams, we can mitigate the impact of rapid changes and fluctuations in predictions.

Recognition and analysis of sports actions in real-time video

DIAB M. Hicham, BOUTICHE Ahmed Imed, LAICHE Nacera, USTHB

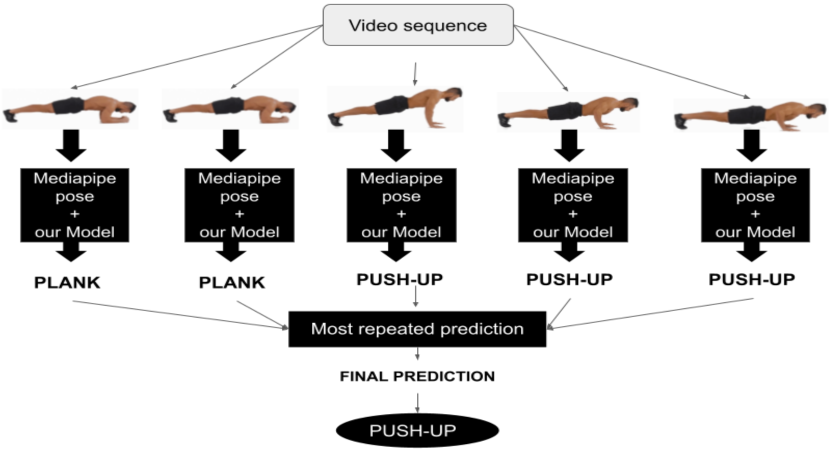

Video sequence
Mediapipe pose + our Model
Mediapipe pose + our Model
Mediapipe pose + our Model
Mediapipe pose + our Model
Mediapipe pose + our Model
PLANK
PLANK
PUSH-UP
PUSH-UP
PUSH-UP
Most repeated prediction
FINAL PREDICTION
PUSH-UP

Figure 4. Queue example illustration

These aspects rely on two measures:

Calculation of Body Joint Angles

Body joint angles provide valuable information about posture, alignment, and movement dynamics, allowing evaluation of exercise techniques. The angle calculation process involves:

- Retrieving the coordinates of three key points representing crucial joints or body parts (Point A, Point B, and Point C).
- Calculating the vectors connecting these points, applying trigonometric functions to determine the angles between the vectors.
- Converting the angles from radians to degrees, and comparing them with minimum and maximum thresholds. These thresholds, experimentally established based on exercise videos produced by professionals, ensure accurate evaluation of exercise execution.

One primary application of these angles is to calculate and increase the number of repetitions of the exercise, as they serve as a measure to track progress and determine the quality of exercise performance.

Calculation of Body Joint Distances

Distances between Mediapipe Pose key points provide valuable insights into movement analysis, body alignment, and balance. The distance calculation process involves:

-Retrieving the coordinates of two key points representing the joints or body parts.
-Using these coordinates to calculate the Euclidean distance between the points.
-Comparing the calculated distance with minimum and maximum thresholds. These thresholds, experimentally established based on exercise videos produced by professionals, ensure accurate evaluation of exercise execution.

Based on the calculated angles and distances, instructions, advice, and repetition count increments are provided to the athlete. It is important to note that the specific angles, distances, thresholds, and instructions vary for each exercise and have been determined experimentally.

Recognition and analysis of sports actions in real-time video

DIAB M. Hicham, BOUTICHE Ahmed Imed, LAICHE Nacera, USTHB

3. Experimentation

In the experimentation phase, the model underwent training and testing to assess its performance. The training process involved monitoring the model's metrics, including loss and accuracy, while the testing phase evaluated its performance on unseen data. The results were analysed using a normalized confusion matrix. Here is an overview of the experimentation process:

Training

During the training process, the model's performance was monitored using appropriate metrics. The evolution of loss, val_loss, accuracy, and val_accuracy was plotted over the training epochs. The graph below represents the metrics' progression.

The plot of figure 5 shows a consistent decrease in both loss and val_loss without fluctuations, indicating effective learning of the model. At the end of training, the loss reached 0.0296, and the val_loss was 0.0538. Simultaneously, the accuracy and val_accuracy steadily increased, reaching values of 0.9903 and 0.9847, respectively.

Testing

To evaluate the model's performance on unknown data, we carried out a test evaluation. The results are as follows:
-Test loss: 0.055
-Test accuracy: 0.984
These results indicate that the model performs well on the test set and generalizes effectively to unknown data.

References

[1] Google. MediaPipe Pose. aout 2020.GitHub, https://github.com/google/mediapipe/blob/master/docs/solutions/pose.md

[2] Brinnae Bent, « InfiniteForm: a synthetic, minimal bias dataset for fitness applications ». Medium, Nov 4, 2021, https://blog.infinity.ai/infiniteform-a-synthetic-minimal-bias-dataset-for-fitness-applications-346e69f35b81

[3] Pismenskova, Marina, et al. « Classification of a Two-Dimensional Pose Using a Human Skeleton ». MATEC Web of Conferences, vol. 132, 2017, p. 05016. https://www.matec-conferences.org/articles/matecconf/abs/2017/46/matecconf_dts2017_05016/matecconf_dts2017_05016.html

[4] Taha Anwar, « Introduction to Video Classification and Human Activity Recognition ». LearnOpenCV, March 8, 2021 https://learnopencv.com/introduction-to-video-classification-and-human-activity-recognition

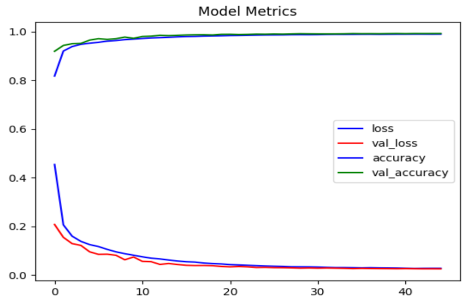


Figure 5. Evolution of the model's metrics

Towards a New Data Representation: GANs for Medical Images Segmentation.

Kenza Aoucher, Ahmed Sif Benmessaoud, Khellaf-Haned H. Faiza, Dahmane A., USTHB

Abstract

Lung cancer remains the leading cause of cancer mortality worldwide. While medical imaging enables earlier diagnosis, precise segmentation of pulmonary nodules from CT scans is critical yet challenging. This work investigates an innovative approach of harnessing generative adversarial networks (GANs) for few-shot learning of lung tumor segmentation models. We hypothesize that GANs encode semantic information when generating realistic medical images. By inverting target CTs through the GAN generator, we extract spatial activation maps to provide localization cues for segmentation with limited examples. Experiments demonstrate promising performance a model trained on just 1 CT segmented lungs scored 10% of traditional full supervision requiring thousands of images. Training on 8 examples matched full supervision. This methodology highlights the potential of leveraging GANs for representation learning and few-shot segmentation in medical imaging.

1. Introduction

Lung cancer has the highest mortality rate of all cancers worldwide, resulting in over 1.7 million deaths per year. Early detection is critical, and computed tomography (CT) screening enables it by identifying malignant pulmonary nodules. However, precise segmentation of lungs and nodules from CTs is challenging and critical for accurate diagnosis and treatment planning. In recent years, artificial intelligence (AI) and computer-aided diagnosis (CAD) systems have emerged to automate lung cancer screening from CT scans using deep learning. Convolutional neural networks can be trained to segment lung regions and detect suspicious nodules in CTs (Figure 1).

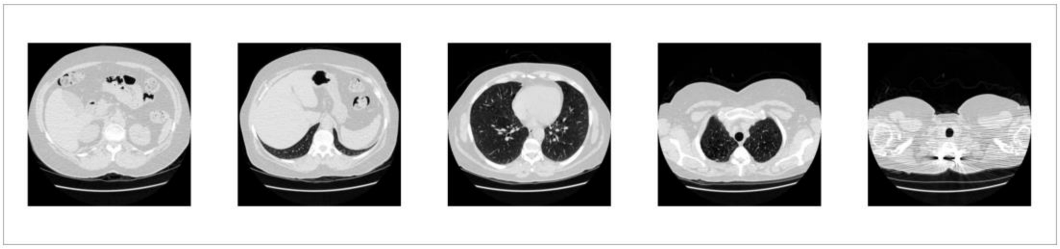


Figure 1. Example of a lung CT scan

2. Method

Our work comprise of two main contributions: the few-shot segmentation pipeline and the computation of highly accurate semantic tumor masks.

2.1 Accurate tumor masks

Experts' annotations in medical imaging datasets exhibit high variability, thus directly combining them reduces segmentation performance. In this work we developed a method to consolidate variable expert masks from the LIDC [1] dataset into precise lung tumor ground truth (Figure 2).

Towards a New Data Representation: GANs for Medical Images Segmentation.

Kenza Aoucher, Ahmed Sif Benmessaoud, Khellaf-Haned H. Faiza, Dahmane A., USTHB

The resulting masks allow for more effective training and evaluation of segmentation models. Tumor center points are expanded into full masks using adaptive region growing and morphology. Multiple expert masks alongside the region growing ones are consolidated into consensus ground truth via voting to mark pixels labeled tumor by over 50% of experts. This captures common tumor regions while avoiding excessive false positives or negatives.

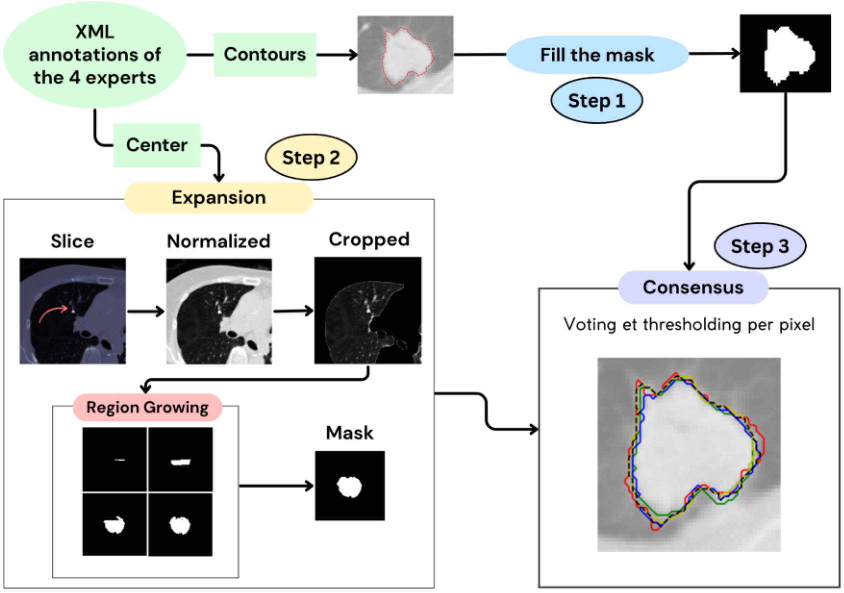


Figure 2. Tumor masks computation methodology

2.2 Few-shot segmentation

Deep learning has driven advances in medical image segmentation, but most methods are supervised thus require very large datasets of pixel-level annotations. Such data is expensive and time-consuming to obtain, especially in the medical field. We propose an innovative approach to overcome these limitations by exploiting generative adversarial networks (GANs) [2] for few-shot semantic segmentation.

GANs are deep neural networks that contain a generator and discriminator that are trained in an adversarial manner (one against the other). Our key hypothesis, as first proposed by Nontawat et al.[3], is that GANs learn to encode the semantic information when synthesizing realistic images. In other words, the generator needs to understand what makes an image an image to be able to generate it. The GAN's feature maps are then used as a pixel-wise representation of the images to train a segmentation network. This allows learning from very few examples while maintaining performance similar to traditional networks trained on massive datasets. Our methodology is exposed in Figure 3.

Towards a New Data Representation: GANs for Medical Images Segmentation.

Kenza Aoucher, Ahmed Sif Benmessaoud, Khellaf-Haned H. Faiza, Dahmane A., USTHB

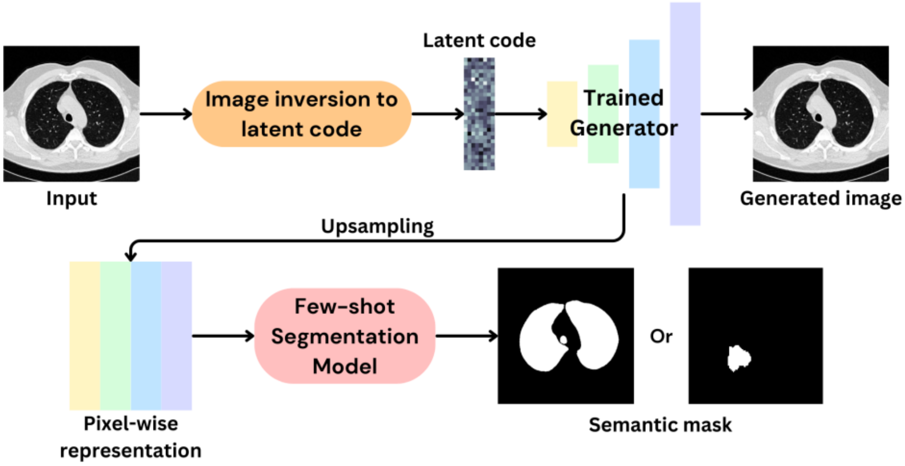

Latent code
Input
Image inversion to latent code
Trained Generator
Generated image
Upsampling
Pixel-wise representation
Few-shot Segmentation Model
Or
Semantic mask

Figure 3. Few-shot segmentation methodology

After careful data preprocessing and creation of reliable semantic masks, advanced GAN architectures are leveraged along with inversion to obtain customizable feature representations of target images. These representations capture details at different levels, from simple textures to complex components. By inverting images to latent vectors, feeding them into the GAN, and concatenating the resulting multi-resolution feature maps, distinctive segmentations are produced with little data. Our work combines the generalization capabilities and component understanding of GANs with the precision of specialized segmentation networks. Specifically, we trained a StyleGAN V3 [4] on over 1 million iterations of CT scans from the Lung Nodule Analysis dataset [5] using 2 NVIDIA GT1070 GPUs over several weeks. Quantitative assessment demonstrated StyleGAN's superior image quality and resolution compared to other models. To invert images, we adapted several optimization algorithms from literature including the W, W+ and PTI methods. Feature maps were then extracted from StyleGAN generator's convolutional layers to serve as pixel-wise representations for few-shot segmentation.

3. Results

Finally, we trained a segmentation model on the GAN-derived representations in a K-shot learning approach and evaluated against full supervision baselines. Experiments show that the 1-shot achieved 90\% of the full-shot performance (Table 1), and the 8-shot demonstrated very promising segmentation of lungs, competing with the traditional fully supervised methods (Figure 4). Tumor segmentation proved more challenging, with few-shot models reaching within 10-15% of full supervision with only 1 training example. Our methodology illuminates the potential of leveraging GANs for representation learning and few-shot segmentation in medical imaging and represents a step toward reducing dependence on large labeled datasets in medical imaging.

Visual Computing MAGAZiNE

Towards a New Data Representation: GANs for Medical Images Segmentation.

Kenza Aoucher, Ahmed Sif Benmessaoud, Khellaf-Haned H. Faiza, Dahmane A., USTHB

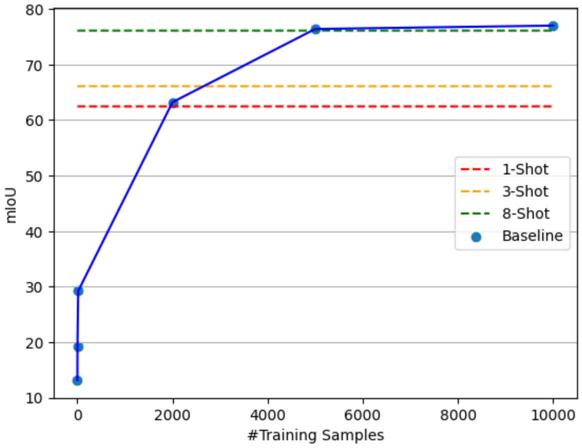
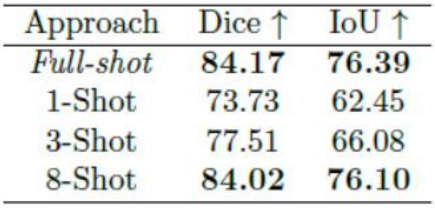


Figure 4. Performance (mIoU) of our lung segmentation model, comparing the full-shot (blue curve) with the few-shots (dotted lines)



Approach	Dice ↑	IoU ↑
Full-shot	**84.17**	**76.39**
1-Shot	73.73	62.45
3-Shot	77.51	66.08
8-Shot	**84.02**	**76.10**

Table 1. Few-shot lung segmentation results

References

[1] Armato III, Samuel G. and McLennan et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans, 2004, National Cancer Institute, url = {wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI.

[2] Ian J. Goodfellow and Jean Pouget-Abadie et al. Generative Adversarial Networks. 2014, arXiv. 1406.2661.

[3] Nontawat Tritrong and Pitchaporn Rewatbowornwong and Supasorn Suwajanakorn . Repurposing GANs for One-shot Semantic Part Segmentation. 2021. arXiv. 2103.04379.

[4] Tero Karras and Miika Aittala and Samuli Laine and Erik Härkönen and Janne Hellsten and Jaakko Lehtinen and Timo Aila. Alias-Free Generative Adversarial Networks. 2021, arXiv. 2106.12423.

[5] Setio, Arnaud Arindra and Ciompi, Francesco et al. Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks. Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2017.

Visual Computing MAGAZiNE

Melanoma identification using deep learning.

Hanafi Yasmine, Laib Wissal, Khellaf-Haned H. Faiza, Dahmane A., USTHB

Abstract

Melanoma is considered one of the most fatal cancer in the world, this form of skin cancer may spread to other parts of the body in case that it has not been diagnosed in an early stage. Thus, the medical field has known a great evolution with the use of automated diagnosis systems that can help doctors and even normal people to determine a certain kind of disease, particularly in deep learning, have provided new opportunities for early detection and diagnosis of melanoma. By utilizing sophisticated algorithms and machine learning techniques, the application developed for melanoma diagnosis aims to improve the accuracy and efficiency of

detecting suspicious skin lesions. By segmenting and classifying these lesions, the mobile application assists in providing timely and reliable information, empowering individuals to seek further medical evaluation and potentially save lives.

1. Introduction

Melanoma, a type of skin cancer, is a serious condition that requires attention (Figure 1). Early detection plays a crucial role in improving patient outcomes. However, accurately identifying and segmenting melanoma lesions from medical images can be challenging. In recent years, artificial intelligence (AI) and computer-aided diagnosis (CAD) systems have emerged as promising tools for automating melanoma screening and diagnosis [1]. Deep learning techniques, such as convolutional neural networks, can be trained to accurately segment and detect suspicious melanomas. These advancements in AI-based technologies have the potential to enhance the accuracy and efficiency of melanoma diagnosis and treatment planning.

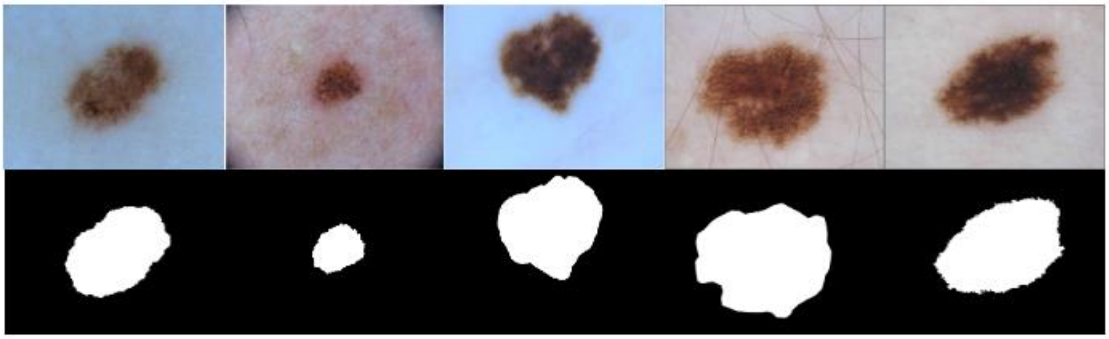


Figure 1. Example of skin cancer images and ground truth.

2. Method

The project's objective is to develop an advanced model for predicting melanoma. Creating such an application involves a series of vital steps. These steps include collecting and pre-processing relevant data, developing segmentation and classification models, training and evaluating these models, and finally integrating the AI model with the application. Through this comprehensive process, the aim is to build a robust and effective tool that can accurately predict and identify melanoma, enhancing early detection and improving patient outcomes.

Melanoma identification using deep learning.

Hanafi Yasmine, Laib Wissal, Khellaf-Haned H. Faiza, Dahmane A., USTHB

2.1 Data collection and pre-processing

The first crucial step in this project is to gather two types of data:
For segmentation, a diverse and representative dataset comprising photos of melanoma lesions along with corresponding ground truth data is necessary. The ground truth data provides accurate outlines and labels for the lesions.

For classification, a dataset of melanoma images representing both benign and malignant cases is required for the classification aspect of the deep learning model. This comprehensive dataset allows for effective training of the model and enables accurate predictions when using the application.
To enhance the dataset and improve the performance of the deep learning model, various techniques for data augmentation were employed (Figure 2). Data augmentation involves applying transformations to the existing images, such as rotation, flipping, scaling, and adding noise. These techniques can create additional diverse examples, expanding the dataset and helping the model generalize better to unseen.

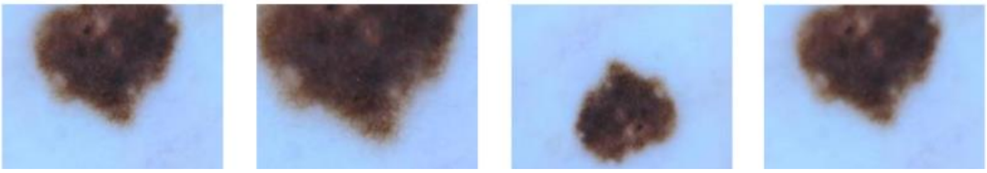


Figure 2. Example of data augmentation operations;

2.2 Image Segmentation

Image segmentation plays a crucial role in the accurate analysis of melanoma cancer. By segmenting the melanoma lesions from medical images, such as dermoscopic images or histopathological slides, the boundaries and characteristics of the cancerous tissue can be precisely delineated, aiding in diagnosis, treatment planning, and monitoring of the disease.

The U-Net architecture, widely known for its effectiveness in medical image segmentation[2], was implemented for the segmentation task in this project (Figure 3). This model has shown great success in accurately delineating melanoma lesions. By incorporating U-Net into the application, the system can reliably identify and analyze melanoma, contributing to improved diagnostics and treatment for patients.

Melanoma identification using deep learning.

Hanafi Yasmine, Laib Wissal, Khellaf-Haned H. Faiza, Dahmane A., USTHB

2.3 Skin lesion classification

Skin lesion classification involves several essential steps to achieve accurate and reliable results. These steps are crucial in the melanoma diagnosis process, as they assign a specific class to each detected lesion. These steps include:
-Data preprocessing: Using Keras ImageDataGenerator class for performing data augmentation on the database to increase the diversity of training data.
-Model selection: Choosing an appropriate classification model is crucial for obtaining good results. We utilized four pre-trained models (ResNet50, Inception V2, InceptionResNet V2, and EfficientNet) to achieve the best possible outcome.
-Model training: Training involves adjusting the model's parameters based on the training data to minimize prediction error. This process may require multiple iterations to achieve optimal performance.
-Model evaluation: The model is evaluated using independent test data. Evaluation metrics such as accuracy, precision, recall, and F-score are calculated to measure the model's performance.

Multiple models were trained and evaluated but ResNet50 was selected as the preferred model for classifying melanoma lesions. The superior performance of ResNet50 ensures reliable and precise predictions, thereby enhancing the overall effectiveness of the application.

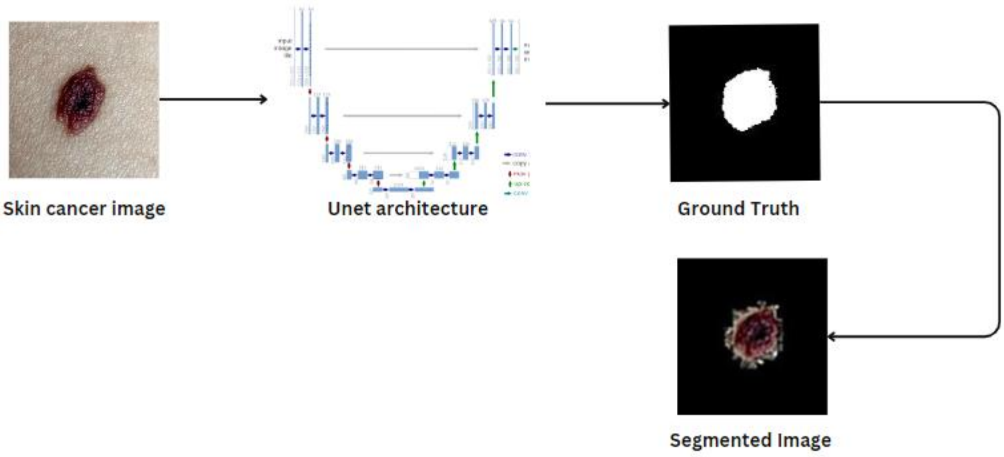


Figure 3. Segmentation methodology.

Melanoma identification using deep learning.

Hanafi Yasmine, Laib Wissal, Khellaf-Haned H. Faiza, Dahmane A., USTHB

3 Results

After rigorous training and evaluation, excellent results were achieved. The segmentation model attained an accuracy of 89 percent, ensuring precise delineation of melanoma lesions (Figure 4, table1). Additionally, the classification model achieved an impressive accuracy of 93 percent, accurately identifying whether the lesions were benign or malignant (Figure 5, Table 2). These high accuracies demonstrate the effectiveness of the models in accurately diagnosing melanoma, contributing to the overall success of the app in providing reliable predictions and aiding in early detection and treatment.

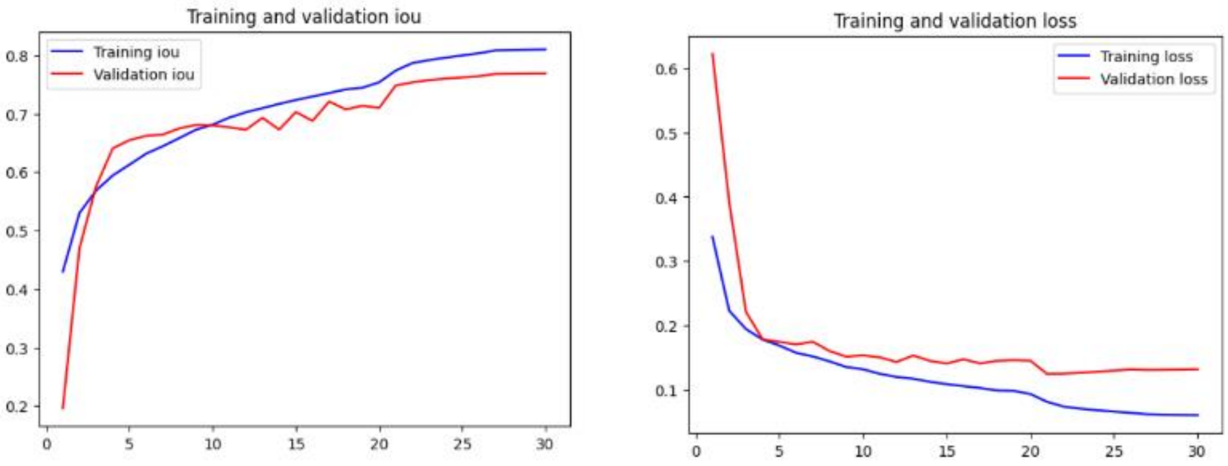
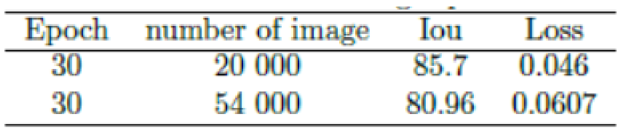


Figure 4. Segmentation results.



Epoch	number of image	Iou	Loss
30	20 000	85.7	0.046
30	54 000	80.96	0.0607

Table 1. The training report for segmentation.

Melanoma identification using deep learning.

Hanafi Yasmine, Laib Wissal, Khellaf-Haned H. Faiza, Dahmane A., USTHB

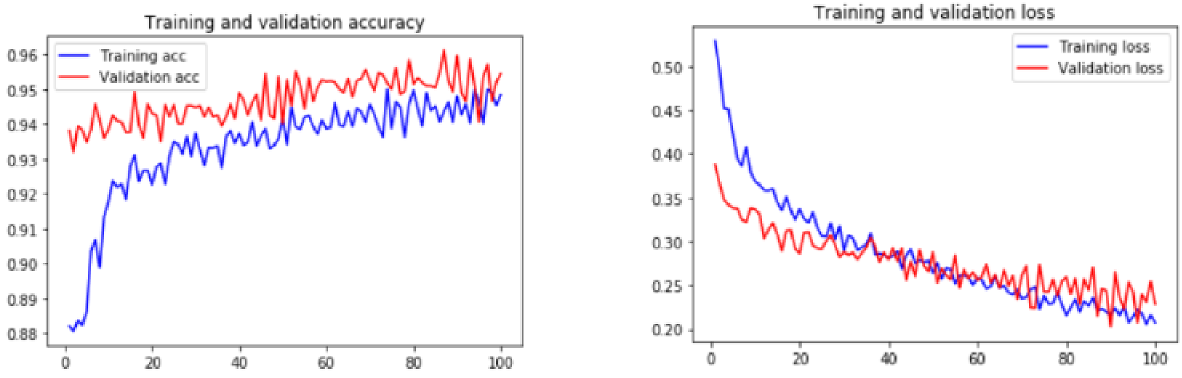
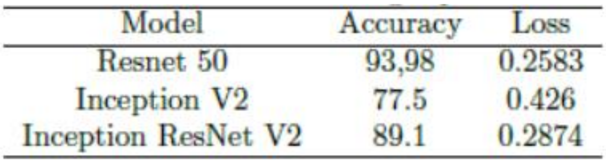


Figure 5. Classification results.



Model	Accuracy	Loss
Resnet 50	93,98	0.2583
Inception V2	77.5	0.426
Inception ResNet V2	89.1	0.2874

Table 2. Training report for classification.

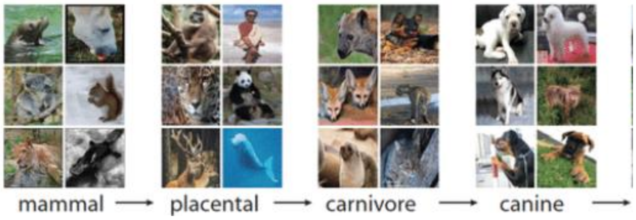
References

[1] J. Daghrir, L. Tlig, M. Bouchouicha, and M. Sayadi, \Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach," in International Conference on Advanced Technologies for Signal and Image Processing, (Sfax, Tunisia), Sept. 2020.

[2] W. Baccouch, S. Oueslati, B. Solaiman, and S. Labidi, \A comparative study of CNN and U-Net performance for automatic segmentation of medical images: application to cardiac MRI," in CENTERIS - International Conference on ENTERprise Information Systems., vol. 219, (Lisbonne, Portugal), pp. 1089-1096, Nov. 2022.

The ImageNet Dataset designed for use in visual object recognition.

Slimane Larabi, USTHB



1. Creation of ImageNet.

Fei-Fei Li was one of the researchers who played an important role in the development of the ImageNet dataset, which has been instrumental in advancing the field of deep learning and computer vision.
She claimed: "The paradigm shift of the ImageNet thinking is that while a lot of people are paying attention to models, let's pay attention to data. Data will redefine how we think about models." [1,2]

In 2009 ImageNet is presented for the first time as a poster at the Conference on Computer Vision and Pattern Recognition (CVPR) in Florida.
ImageNet is a large dataset of annotated photographs intended for computer vision research.
The goal of developing the dataset was to provide a resource to promote the research and development of improved methods for computer vision.

Based on statistics about the dataset recorded on the ImageNet homepage, there are a little more than 14 million images in the dataset, more than 21 thousand groups or classes (synsets), and more than 1 million images that have bounding box annotations (e.g. boxes around identified objects in the images). The photographs were annotated by humans using crowdsourcing platforms such as Amazon's Mechanical Turk [1,2].

2. ImageNet and Deep Learning

Alex Krizhevsky, et al. from the University of Toronto in their 2012 paper titled "ImageNet Classification with Deep Convolutional Neural Networks" developed a convolutional neural network that achieved top results on the ILSVRC-2010 and ILSVRC-2012 image classification tasks.

These results sparked interest in deep learning in computer vision: "we trained one of the largest convolutional neural networks to date on the subsets of ImageNet used in the ILSVRC-2010 and ILSVRC-2012 competitions and achieved by far the best results ever reported on these datasets." [3]

References

[1] J. https://www.image-net.org/about.php
[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, ImageNet: A Large-Scale Hierarchical Image Database. IEEE Computer Vision and Pattern Recognition (CVPR), 2009
[3] Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E. ImageNet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems (NIPS), 2012

Visual Computing Magazine

Vol 1, Issue 2, 2023

Content

Preface	Page 02
Road Object Detection and distance estimation using depth monocular model	Page 03
Classification of Suspected Pulmonary Nodules Based on AI Approaches	Page 08
Object Recognition System on a Tactile Device for Visually Impaired	Page 12
Recognition and analysis of sports actions in real-time video	Page 19
Towards a New Data Representation: GANs for Medical Images Segmentation	Page 24
Melanoma identification using deep learning	Page 28
The ImageNet Dataset designed for use in visual object recognition	Page 33

Call for papers

Authors working in Visual Computing are invited to submit papers (3-6 pages) to the magazine, email: vcm@usthb.dz

Visual Computing Magazine