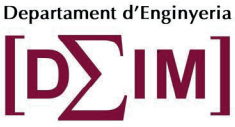
7TH URV DOCTORAL WORKSHOP IN COMPUTER SCIENCE AND MATHEMATICS

Edited by
Mohamed Abdel-Nasser, Oriol Farràs,
Domènec Puig, Hatem A. Rashwan

UNIVERSITAT ROVIRA i VIRGILI



Title: 7th URV Doctoral Workshop in Computer Science and Mathematics
Editors: Mohamed Abdel-Nasser, Oriol Farràs, Domènec Puig, Hatem A. Rashwan
March 2022

Universitat Rovira i Virgili
C/ de l'Escorxador, s/n
43003 – Tarragona
Catalunya (Spain)

ISBN: 978-84-1365-033-3
DOI: 10.17345/9788413650333

CONTENT

Preface
Mohamed Abdel-Nasser, Oriol Farràs, Domènec Puig, and Hatem A. Rashwan

DeepKey: Watermarking Deep Learning Models
Najeeb Jebreel

End User's Side Explanations of DL Models' Predictions
Rami Haffar

Distance and Size Calculation of the detected objects on Floor from robot using Bounding Box
Aditya Singh

Quality of Life Analysis of Dependent People using Multiple Linear Regression Model
Gaurav Kumar Yadav

Dynamic update of Fuzzy Random Forests to improve classification of Diabetic Retinopathy
Jordi Pascual-Fontanilles

Monocular Depth Estimation with Self-supervised Graph Convolutional Network
Armin Masoumian

Contributions to GDPR compliance by means of Smart Contracts
Cristòfol Daudén-Esmel

Deep learning-Based Approach for Retinal Lesions Segmentation in Eye Fundus Images
Moahammed Yousef Salem Ali

Fundus Image Quality Assessment Based on Deep Autoencoder Networks
Saif Khalid

Radiomics-based computer-aided diagnosis system for prostate cancer classification in MRI images
Eddardaa Ben Loussaief

Breast Tumor Segmentation in Ultrasound Image using Deep Learning Techniques
Nadeem Issam Zaidkilani

Optimizing the first convolutional layer
João Paulo Schwarz Schüler

Efficient Data Augmentation Techniques for Lesion Detection in Breast Tomosynthesis Images Using Deep Learning Models
Loay Hassan

Road Damage Detection Using Yolov5
Ammar Mohammed Okran

PREFACE

This book contains the abstracts of the works presented in the 7th Doctoral Workshop in Computer Science and Mathematics - DCSM 2022. It was celebrated in Universitat Rovira i Virgili (URV), Campus Sescelades, Tarragona, on March 31, 2022. The aim of this workshop is to promote the dissemination of ideas, methods, and results developed by the students of the PhD program in Computer Science and Mathematics from URV. It has been jointly organized by the research group of Intelligent Robotics and Computer Vision (IRCV) and the Doctoral Program on Computer Science and Mathematics of Security of URV.

The editors and organizers invite you to contact the authors for more detailed explanations and we encourage you to send them your suggestions and comments that may certainly help them in the next steps of their PhD thesis. We thank all the participants and, especially, the students that presented their work in this DCSM workshop. Finally, we also want to thank Universitat Rovira i Virgili, the Departament d'Enginyeria Informàtica i Matemàtiques (DEIM), and the Escola Tècnica Superior d'Enginyeria (ETSE) for their support.

Mohamed Abdel-Nasser, Oriol Farràs, Domènec Puig, and Hatem A. Rashwan

DeepKey: Watermarking Deep Learning Models

Najeeb Jebreel *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Catalonia
najeeb.jebreel@urv.cat

Abstract. Many organizations devote significant resources to building high-accuracy deep learning (DL) models. Thus, they have a great interest in protecting their trained models from being stolen or misused. Embedding watermarks (WMs) in DL models is a useful means to protect their owners' intellectual property (IP). This paper proposes *DeepKey*, a novel watermarking framework for DL models. We leverage multi-task learning (MTL) to learn the original classification and watermarking tasks jointly. Empirical results show that *DeepKey* can preserve the utility of the original task and embed a robust WM.

Keywords: Deep learning models; Ownership; Intellectual property; Watermarking.

1 Introduction

Deep learning models' owners, such as technology companies, devote significant resources to train their models on vast amounts of proprietary training data, whose collection also implies a significant effort [1]. Thus, they seek compensation for the incurred costs by reaping profits from commercial exploitation [3]. Due to the competitive nature of the technology market, a stolen or misused model is clearly detrimental to its owner on both economic and competitive terms. Therefore, legitimate owners need a robust and reliable way to prove their ownership of DL models in order to protect their intellectual property (IP). Embedding watermarks (WMs) in DL models is a useful means to protect their owners' intellectual property (IP) [2].

We propose *DeepKey*, a novel watermarking framework that allows owners to embed reliable and robust digital WMs in their DL models. Extensive experiments show that *DeepKey* can successfully embed robust WMs with reliable detection accuracy while preserving the accuracy of the original task. The remainder of the paper is organized as follows. Section 2 presents an overview of our framework. Section 3 reports the experimental results. Finally, Section 4 gathers conclusions and proposes several lines of future research.

2 The DeepKey framework

The key idea of our framework is to perform two tasks at the same time: the original classification task T_{org} and the watermarking task T_{wm}. Fig. 1 shows the global workflow of *DeepKey*.

Watermark embedding. *DeepKey* takes four main inputs in the WM embedding phase: the target model (pre-trained or from scratch), the original data set, the owner's WM carrier set, and the owner's information string. The output is the marked model, corresponding private model, and the owner's signature. First, the WM carrier set samples are signed using the owner's signature. After that, the signed WM carrier set is combined

* PhD advisor: Josep Domingo-Ferrer

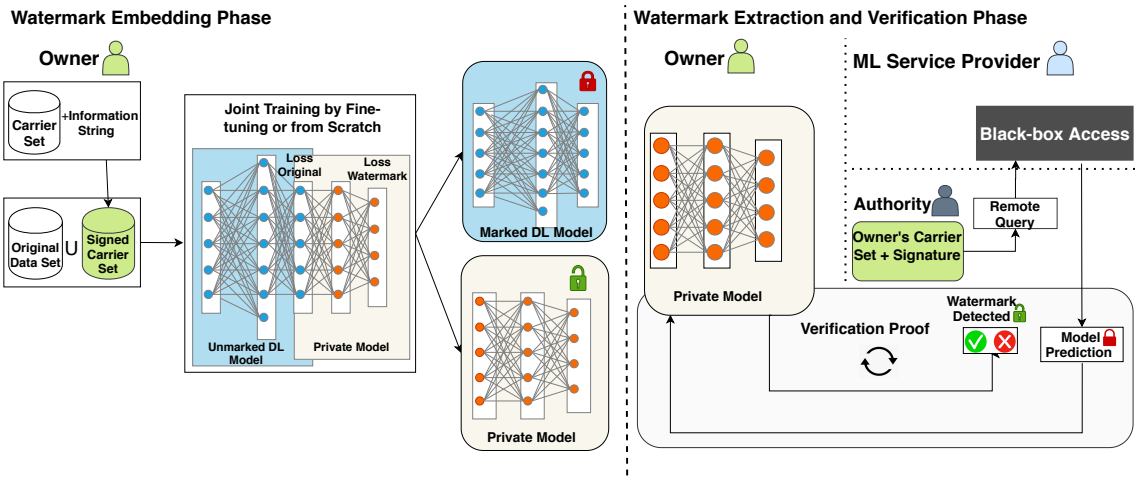


Fig. 1: *DeepKey* global workflow.

with the original data set, and they are used to fine-tune (or train) the targeted model. Finally, the private model takes the final predictions of the original model as inputs and outputs the position of the owner's signature on the WM sample. We leverage MTL to train the two models jointly.

Watermark extraction and verification. To extract and verify the ownership of a remote black-box DL model, the owner first delivers the WM carrier set and her signature to the authority. She also tells the authority about the methodology used to sign the WM samples and the predefined positions where the WM may be placed. Next, the authority (*i.e.*, the *verifier*) randomly chooses a sample from the carrier set, puts the signature in a random position, queries the suspicious remote DL model and sends the model's predictions to the owner. The owner (*i.e.*, the *prover*) takes the predictions, passes them to her private model, and tells the authority the position of her signature on the image. The authority repeats the proof as many times as she desires. After that, the owner's answer accuracy is evaluated according to a minimum threshold. If the owner surpasses the threshold, her ownership is regarded as proven by the authority.

3 Experimental results

Original and watermark tasks data sets and models. We used the CIFAR10 data set for the original task while we used STL10 as a WM carrier set. We used ResNet18 and VGG16 DL models for the original task while we used a simple DL model (with 496 learnable parameters) as a private model. Fig. 2 shows some examples of signed carrier set images and their corresponding labels.

We used *accuracy* as a performance metric for the original task and the WM task. We set the required threshold $T = 90\%$ to prove model ownership. In the following, we evaluate the fidelity, reliability and integrity of *DeepKey*. Also, we assess its robustness against two types of attacks: *fine-tuning* [4] and *model compression* [5].

Fidelity and reliability. Embedding the WM should not decrease the accuracy of the marked model on the original task. As shown in Tab. 1, *DeepKey* did not degrade the accuracy of the original task and successfully embedded the watermark. This is thanks to the joint training, which simultaneously minimizes the loss for the original task and the WM task. Also, it shows that legitimate owners were able to reliably prove their ownership with accuracy greater than 90%.

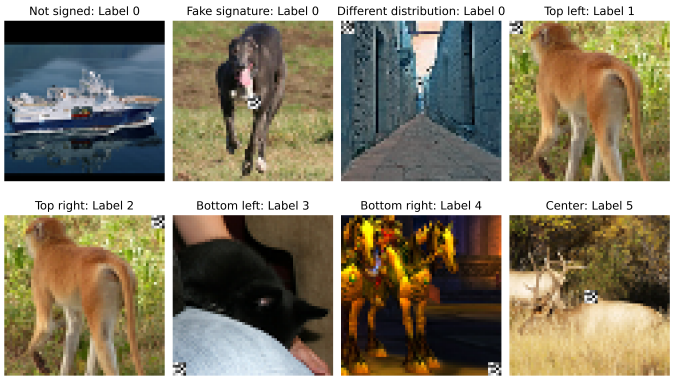


Fig. 2: Examples of signed STL10 carrier set images employed with the CIFAR10 data set.

Integrity. *DeepKey* yields low WM accuracy detection with unmarked models, and thus it does not falsely claim ownership of models owned by a third party. In our experiments, there were 6 classes for the watermarking task. Looking at Tab. 2, the accuracy of falsely claimed ownership of unmarked models is not far from guessing 1 out of 6 numbers randomly, which equals approximately 16%.

Table 1: Accuracy of the original and the WM tasks

Benchmark	**Unmarked model accuracy %**	**Marked model accuracy %**		**WM detection accuracy %**	
		By finetuning (30 epochs)	**From scratch (250 epochs)**	**By finetuning**	**From scratch**
CIFAR10-ResNet18	91.96	92.07	92.53	99.96	99.97
CIFAR10-VGG16	90.59	90.52	91.74	99.68	99.89

Table 2: Integrity results with unmarked models. Each private model was tested with two different unmarked models: one model has the same topology as its corresponding marked model, the other one has a different topology. The last four columns show the accuracy detection obtained with the unmarked models.

Dataset	**DL model**	**Watermark detection accuracy with marked models%**	**Watermark detection accuracy with unmarked models %**			
			Same topology	Accuracy	Different topology	accuracy
CIFAR10	**ResNet18**	99.97%	ResNet18	18.92%	VGG16	19.80%
CIFAR10	**VGG16**	99.89%	VGG16	7.92%	ResNet18	12.32%

Robustness. Tab. 3 show that *DeepKey* was robust to the fine-tuning attacks for a number of fine-tuning epochs ranges from 50 to 200. Fig. 3 shows that *DeepKey* is robust against model compression, and the accuracy of the WM remains above the threshold $T = 90\%$ as long as the marked model is still useful for the original task.

4 Conclusion

We have presented *DeepKey*, a novel watermarking framework that enables DL model owners to embed robust watermarks in their models while preserving the accuracy of the main task. As future work, we plan to extend *DeepKey* to watermark federated deep neural networks.

Table 3: Robustness to model fine-tuning

Benchmark	CIFAR10-ResNet18			CIFAR10-VGG16		
# of epochs	50	100	200	50	100	200
Marked model accuracy %	92.40	92.30	92.47	91.31	91.64	91.69
WM detection accuracy %	98.19	98.05	99.12	97.20	94.72	96.67

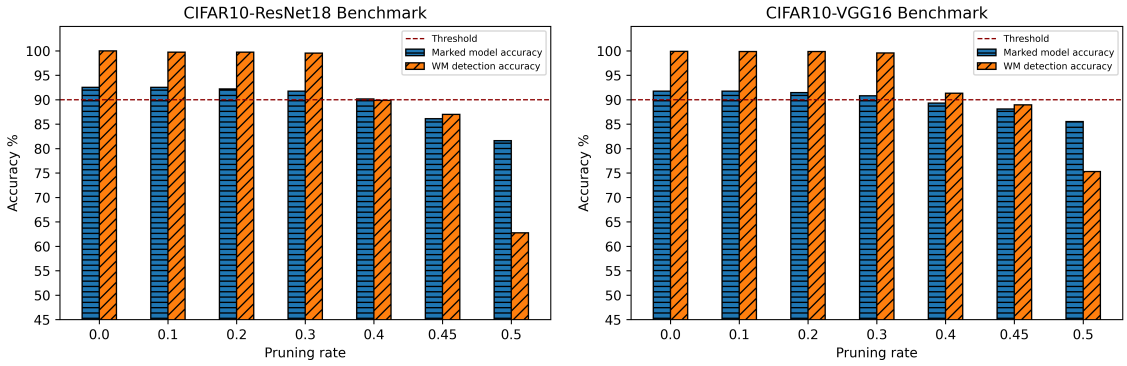


Fig. 3: Robustness against model compression

Acknowledgement. This research was funded by the European Commission (projects H2020-871042 "SoBigData++" and H2020-101006879 "MobiDataLab"), the Government of Catalonia (FI grant to N. Jebreel).

References

[1] Adiwardana, Daniel and Luong, Minh-Thang and others. Towards a human-like open-domain chatbot. *arXiv preprint arXiv:2001.09977.*, 2020.

[2] Jebreel, Najeeb Moharram and Domingo-Ferrer, Josep and Sánchez, David and Blanco-Justicia, Alberto. KeyNet: An Asymmetric Key-Style Framework for Watermarking Deep Learning Models. *Applied Sciences.*, Multidisciplinary Digital Publishing Institute, v. 11, no. 3 p. 999, 2021.

[3] Ribeiro, Mauro and Grolinger, Katarina and Capretz, Miriam AM. Mlaas: Machine learning as a service. *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA).*, IEEE, pp. 896-902, 2015.

[4] Tajbakhsh, Nima and Shin, Jae Y and Gurudu, Suryakanth R and others. Mlaas: Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging.*, IEEE, v. 35, no. 5, pp. 1299-1312, 2016.

[5] Han, Song and Mao, Huizi and Dally, William J. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. *arXiv preprint arXiv:1510.00149.*, 2015.

End User's Side Explanations of DL Models' Predictions

Rami Haffar *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili Tarragona, Catalonia
rami.haffar@urv.cat

Abstract. Deep learning (DL) models are being used to solve various critical tasks in the past few years. However, those models are ambiguous in terms of how their predictions are made. For the end users to trust those models, the end users should have the ability to generate local explanations of the predictions made by the DL models. In this work, we present a novel approach allowing an end user to locally generate explanations for a DL classification model accessed via a provider's API. We approximate the provider's model with a local surrogate model. We then use the surrogate model's components to locally generate explanations.

Keywords: End-user explanations; Deep learning; Counterfactual explanations.

1 Introduction

Building highly accurate Deep learning (DL) classification models requires a large amount of training data, whose collection and labeling involve a significant effort. Therefore, small businesses and ordinary users, who cannot afford this effort, resort to big technology companies that provide paid API access to highly accurate DL models via Machine Learning as a Service (MLaaS) platforms [4]. These end users then query those models with their (small) data and obtain the final classification predictions.

Even though end users are interested in using MLaaS with highly accurate DL models, they may not entirely trust such models due to the lack of transparency of DL predictions. Obtaining explanations alongside predictions helps end users understand why a DL model produces a specific prediction, which increases the trust in the model and contributes to clearer decision-making [1].

We propose a novel approach that allows an end user to locally generate DL model-specific explanations for a DL classification model accessed via a provider's API. The approach consists of two main phases: i) approximating the provider's model by a local surrogate model, using the small portion of data owned by the end user and ii) using the surrogate model to locally

* PhD advisor: Josep Domingo-Ferrer, and David Sánche¡

generate DL model-specific explanations that approximate the explanations obtainable with white-box access to the provider's model. The remainder of the paper is organized as follows. Section 2 presents an overview of our proposed method. Section 3 reports the experimental results. Finally, Section 4 gathers conclusions and proposes several lines of future research.

2 Explaining deep learning classification model predictions on the user's side

The importance of our proposed method lies in allowing users to reliably understand how the providers' models make their predictions and determine whether these predictions are trustworthy.

In the first phase, we need to approximate the provider's model by a local surrogate model having an accuracy as close to that of the provider's model as possible. However, the end user does not own enough data to train such a model. To tackle this challenge, we employ a modified version of the Mixup method [6] to augment the user unlabeled data and obtain more representative training data. Once the user obtains the augmented data, she queries the provider's API to label the data. Afterward, due to the model knowledge from a complex "master" to a simpler "student" model being transferable [3], the end user trains a local surrogate model using the labeled data she recently acquired.

Once the end user obtains the trained local surrogate model she can use it to generate accurate explanations for the predictions of the provider model. Since the surrogate model has almost learned the same decision boundaries as the provider's model, explanations generated using the surrogate's internal components can be expected to accurately approximate the explanations generated using the provider's internal components.

2.1 Generating the explanations.

In our work, we explain the provider's model by generating counterfactual explanations [5] of a specific example. Counterfactual explanations tell us how to change the example's features so that its predicted label also changes. In this way, we can understand how the model makes its predictions and explain individual predictions. We use adversarial training [2] as a means to generate adversarial examples that counterfactually explain the model predictions. In fact, adversarial examples are aimed at fooling the model rather than explaining it, but, in the end, they serve the same purpose as counterfactual examples by slightly changing the features of input examples to modify their predicted labels.

3 Experimental results

Models and data sets. We use the gender classification and MNIST data sets to test the performance of the proposed method on image data. In all the experiments we took the surrogate model to be simpler than the provider's model, which can save training time and at the same time retains most of the provider's model knowledge.

We used the following evaluation metrics to measure the performance of the trained surrogate models and the generated explanations:

- Accuracy: We used this metric to measure and compare the performance of the provider's and the surrogate models.
- Structural Similarity Index Measure (SSIM): We used SSIM to measure the similarity between the explanations provided by the surrogate model and the ones generated by the provider's model for image data.

Accuracy of surrogate model. Table 1 reports the accuracy of the provider's models and the trained surrogate models. We can see that the performance of the surrogate models was nearly equivalent to that of the provider's model.

Table 1. Accuracy of the Provider's model and surrogate model.

Data set	Provider model	Surrogate model
Gender	96.3%	94.47%
MNIST	99.22%	96.1%

Surrogate model explanation. Table 2 reports the average SSIM for the Gender and MNIST validation images. We can see that the surrogate model generates explanations (counterfactual examples) with very high similarity to those generated by the provider's model, which indicates that the surrogate model is properly approximating the provider's model.

Table 2. Similarity between the adversarial examples generated by the surrogate model and those generated by the provider's model on the Gender and MNIST data sets.

Gender	**96.79%**
MNIST	**98.27%**

Figure 1 shows two examples of these visual explanations generated for the Gender and MNIST data sets. By looking at the pixels that caused the prediction to change, we can see that, in general, the explanations generated by the surrogates were consistent with those generated by the provider's model:

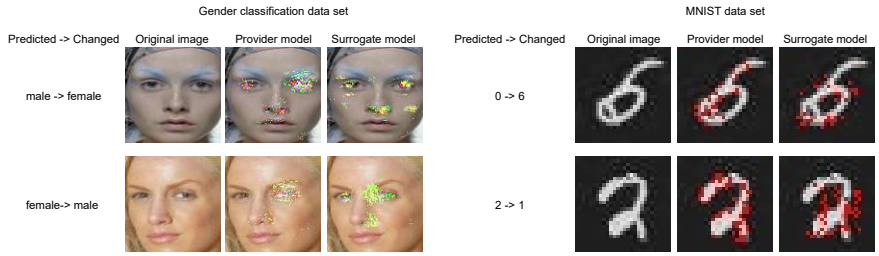


Fig. 1. Visual explanations generated by the surrogate model in comparison with those generated by the provider's models.

4 Conclusion

We have presented a novel approach that enables the end user to locally generate explanations on the predictions of the provider's model. As future work, we plan to test the performance of our approach on other computer vision tasks, such as detection and segmentation, as well as natural language processing.

Acknowledgement. We acknowledge support from the European Commission (projects H2020-871042 "SoBigData++" and H2020-101006879 "MobiDataLab").

References

[1] Blanco-Justicia, A., Domingo-Ferrer, J., Martinez, S., Sanchez, D. Machine learning explainability via microaggregation and shallow decision trees. *Knowledge-Based Systems*, Knowledge-Based Systems, 2020.

[2] Bruna, J., Szegedy, C., Sutskever, I., Goodfellow, I., Zaremba, W., Fergus, R., Erhan, D. Intriguing properties of neural networks. *arXiv preprint* arXiv:1312.6199. 2013.

[3] Hinton, G., Vinyals, O., Dean, J. Distilling the knowledge in a neural network, *arXiv preprint* arXiv:1503.02531, 2015.

[4] Ribeiro M, Grolinger K, Capretz MA. Mlaas: Machine learning as a service. *IEEE 14th International Conference on Machine Learning and Applications (ICMLA)* (pp. 896-902). IEEE. 2015.

[5] Wachter, S., Mittelstadt, B. and Russell, C. Counterfactual explanations without opening the black box: Automated decisions and the GDPR", *Harvard Journal of Law and Technology* 31(2017) 841.

[6] Zhang, H., Cisse, M., Dauphin, Y. N., Lopez-Paz, D., mixup: Beyond empirical risk minimization, *International Conference on Learning Representations*. 2018.

Distance and Size Calculation of the detected objects on Floor from robot using Bounding Box

Aditya Singh *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
aditya.singh@urv.cat

1 Introduction

This work aims to develop a mathematical relation between the position of an object in a two dimensional image plane and three dimensional world space with respect to a robot mounted camera by using object detection bounding box coordinates. In human-centric robot navigation, it is very necessary to make a perception of object distance and position with respect to robot. It uses the object detection information in the form of bounding box by using monocular vision and uses the robot kinematic parameters to establish a mathematical relation between object and robot camera. The position of the object is calculated in two fold: one is by calculating the distance in front direction and other is by doing side positioning.

2 Methodology

The process is tested on a Locobot Robot (PyRobot [1] platform, developed by Trossen Robotics). The robot uses a Intel Realsense camera for vision. The process uses YOLOv3 algorithm for object detection. It uses Manhattan World Assumption [2] for defining the floor as a horizontal plane and in 3D world, all the pixels from the floor lie in a single plane.

2.1 Object Detection and ground object discovery

A YOLOv3 model takes an Image $I(x, y)$ as input and predict the objects present in the image. The output of the object detection is (H_i, W_i, C_i), which are the dimensions of the bounding box for i object. For identifying the ground located objects, bounding box dimensions play a crucial role. As shown in figure 1 and figure 2, the pixel or pixel height (H_t) which corresponds to

* PhD advisors: Prof. Domenec Puig
Prof. G. C. Nandi, IIIT Allahabad.

the camera height world space are taken as reference for the calculation. The peculiarity of this pixel height is its 2D nature i.e. the height of the world points corresponding to this height will not change in image plane of a camera for a 2D motion. The objects, whose lower side of bounding box is below this line is considered as a floor located object. This assumption works well for major cases due to low height of the camera.

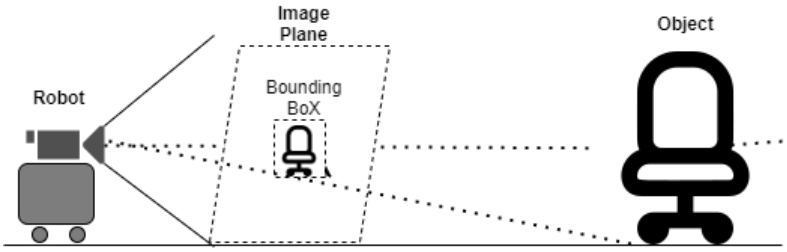


Fig. 1. Side view for relation between camera view and environment.

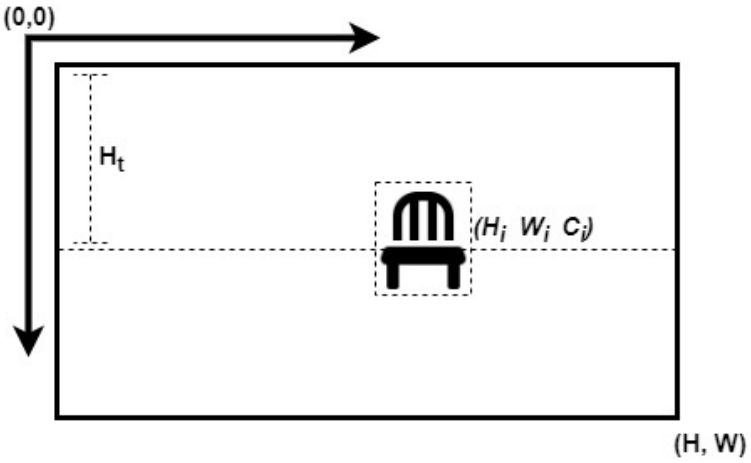


Fig. 2. Image plane information for object detection output.

2.2 Distance Measurement

After floor object identification, bounding box coordinates are used for calculating its position on ground. The height of the lower side of bounding box is $C_i + H_i/2$ and represented by 'H_l' (all the distances are measured with respect to origin of the image plane (0,0) as shown in figure 2). The horizontal field of view (FoV) of camera is 'θ_h', vertical FoV is 'θ_v', the height of the camera

is 'h', the angle made by lower side of object bounding box and lower side of 'θ_v' on camera lens is 'θ_o' and angle made by 'H_t' and lower side of 'θ_v' is 'θ_t'. By perpendicular triangle law the distance between the camera and robot is given by,

$$Z = h/\tan(\theta_h - \theta_o) \tag{1}$$

and 'θ_h' is calculated as,

$$\theta_h = \theta_v \times H/(H - H_t) \tag{2}$$

2.3 Object Size Measurement

The distance (Z) of the object from the robot will become the reference for calculating the height, width and its placement in terms of left or right. The projection of the object is considered perpendicular to the floor. This calculation considers the focal distance 'f' of the robot camera and by using focal distance and 'Z' every point of the image plane can be mapped in the real world by using,

$$(X, Y) = Z \times ((x_0, y_0) - (x_i, y_i))/f \tag{3}$$

where, (x_0, y_0) are the coordinates of the centre of the image plane and (X, Y) is the deviation of the point from the center line of sight of the robot camera.

3 Progress

This idea is used for Locobot robot and used for Semantic Mapping of indoor environment. The results are good and the most interesting thing is its light functionality. It can run on any kind of robot processor for calculating distance with objects. In figure 3, distance prediction results are shown for images taken from laboratory environment.

Fig. 3. Results for distance measurement of detected objects

Acknowledgement. The research work is supported by 'Semantic Mapping' grant provided by 'URV Spain'.

References

[1] Murali, Adithyavairavan, Tao Chen, Kalyan Vasudev Alwala, Dhiraj Gandhi, Lerrel Pinto, Saurabh Gupta, and Abhinav Gupta. "Pyrobot: An open-source robotics framework for research and benchmarking." arXiv preprint arXiv:1906.08236 (2019).

[2] Coughlan, James and Yuille, Alan L. The Manhattan world assumption: Regularities in scene statistics which enable Bayesian inference, Advances in Neural Information Processing Systems, 13, 845–851, 2000.

Quality of Life Analysis of Dependent People using Multiple Linear Regression Model

Gaurav Kumar Yadav *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili Tarragona, Spain
gauravkumar.yadav@urv.cat

1 Introduction

The concept of quality of life (QOL) is difficult to operationalize. However, the recent development in this area reports that improved quality of life is a realistic and obtainable goal for everyone, including people with intellectual disabilities (ID). This work aims to analyze the dataset recorded during an interview of an individual, older people, or people with intellectual disabilities. The interviewer asked questions related to the dimensions of QOL. Many research works [1], [2] propose eight dimensions of QOL. These eight dimensions are Emotional Well-being (BE), Interpersonal Relation (IR), Material Well-being (BM), Personal Development (DP), Physical Well-being (BF), Self-determination (AU), Social Inclusion (IS), and Rights (DR). Each dimension has four to six objective questions related to that Dimension. Based on the answers of each dimension of an individual, an interviewer who is a professional gives an index value of QOL. The index value shows the output of the corresponding eight dimensions of the quality of life. We have interviewed a total of twenty-six individuals and built a dataset. We use a multiple linear regression model to analyze the dataset.

2 Methodology

This work is motivated to build a learning machine that can evaluate the quality of life of an individual and, based on evaluation, suggest the possible support in the particular Dimension. Figure 1 reports the complete layout of the work. It starts with the individual's interview and asking questions related to the eight dimensions of QOL. The answer records based on 3 points Likert scale and recorded the converted answers for each individual in every eight dimensions. Furthermore, an expert gives corresponding support index values based on input values. In step two, we recorded the data for each

* PhD advisor: Prof Domenec Puig, Prof G.C.Nandi

individual. We prepare the data for a learning algorithm by pre-processing it. After training, the model predicts the expected support index value for new incoming QOL dimension values. Based on the support index value, we detect the deficiencies in any dimension of QOL, and professionals make an action plan to improve one or more aspects of QOL. This process collectively runs timely and tries to enhance the quality of life of individuals, including the person with intellectual disabilities.

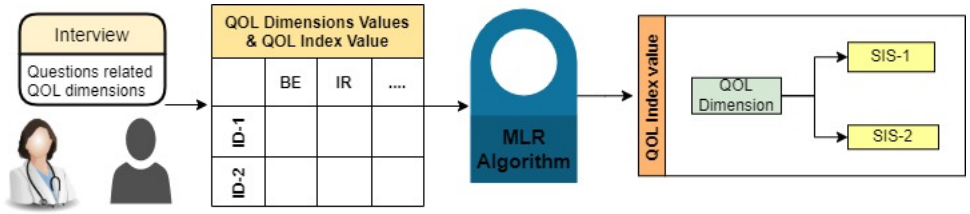


Fig. 1. Complete architecture to evaluate Quality of Life of an Individual

2.1 Dimension of Quality of Life and Support Intensity Scale

The development in this area motivates the dimension from striving to define QOL to focusing its basic dimensions [1]. It shows that QOL is a multidimensional phenomenon than an individual trait. People's QOL is affected by the interaction between personal and environmental factors. Therefore, Its evaluation is based on subjective and objective measures. Recent research [3] depicts the development of a new paradigm that integrates QOL with support (QOLSP).

Dimension of Quality of Life	Area of Scale of Support Intensity Scale
Emotional Well-being	Health and Healthcare, Protection and defence, and Behavioral support need
Interpersonal Relations	Social activities
Material Well-being	Employment activities
Personal Development	Homelife activities, life long learning
Physical Well-being	Health and Healthcare, Exceptional medical need
Self-determination	Protection and defence
Social Inclusion	Community life activities, Social activities
Rights	Protection and defense, Health and Healthcare

Table 1. Dimension of quality of life and area of the support intensity scale.

The multidimensional concept of QOL proposes the various factor to decide the dimension of the QOL. These factors are independence, social participation, and well-being [2]. Based on these factors, dimensions are following shown in Table 1. The support corresponding to these dimensions is also shown in the second column of Table 1. These dimensions encompass every part of a

person's personality. Various indicators show that the area needs to work to improve the quality of life dimension.

2.2 Multiple Linear Regression Model

MLR (Multiple Linear Regression) is a popular regression algorithm for solving scenarios with multiple input attributes. QOLSP contains eight input dimensions, each with its own support index, and MLR predicts the corresponding support index value based on these eight input dimensions of an individual's QOL. This algorithm is implemented using the equation below.

$$Y_i = W_0 + {W_1}^T X_{i1} + {W_2}^T X_{i2} + {W_3}^T X_{i3} + ... + {W_p}^T X_{ip} + \epsilon \qquad (1)$$

Where Y_i represents the support index, X_i represents the QOL dimensions, W_0 represents the bias, and W_p represents the slope coefficients for each QOL dimension. The model error is shown by *epsilon*. We divided the dataset into 80 percent for training and 20 percent for testing.

3 Progress

This study utilizes a machine-learning system to predict the corresponding support index value. Furthermore, with the assistance of dimension specialists, we must create an action plan that corresponds to the support value and provide it to the individual with a matching action sheet. We calculated train case accuracy and subsequently test case accuracy in the form of mean absolute error, root mean square error, and R^2 score after training the model with the training dataset, as shown in table 2.

Evaluation Matrices	**Train Case**	**Test Case**
Absolute Mean Error	0.490262	1.350473
Root mean square error	0.635070	1.472938
R^2 score	0.998192	0.991830

Table 2. Evaluation matrices for quality of life evaluation

Acknowledgement. This project is supported by URV Tarragona and fundacio Ave maria foundation.

References

[1] R. L. Schalock, "The concept of quality of life: what we know and do not know," *Journal of intellectual disability research*, vol. 48, no. 3, pp. 203–216, 2004.

[2] J. Van Loon, "Un sistema de apoyos centrado en la persona. mejoras en la calidad de vida a través de los apoyos," 2013.

[3] L. E. Gómez Sánchez, R. L. Schalock, M. Á. Verdugo Alonso *et al.*, "A new paradigm in the field of intellectual and developmental disabilities: Characteristics and evaluation," *Psicothema*, 2021.

Dynamic update of Fuzzy Random Forests to improve classification of Diabetic Retinopathy

Jordi Pascual-Fontanilles *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
jordi.pascual@urv.cat

1 Introduction

We want to address the problem of Diabetic Retinopathy (DR) classification. As a consequence of diabetes, the blood vessels of the eye may break and generate small blood spots, hemorrhages and exudates. These lesions produce vision loss and may even cause blindness if they are not detected and treated at an early stage.

In this PhD thesis, the goal is to improve a DR classification model based on a Fuzzy Random Forest (FRF) used in the Retiprogram system [2] [3]. It uses the clinical data of the patients to assess their risk of developing DR. This model is being tested by a group of ophthalmologists at Hospital Sant Joan de Reus. The general results are good (with a sensitivity and a specificity over 75%), but there are still many miss-classifications. Errors are mainly due to the inherent ambiguity of the training examples (very similar patients can belong to different classes) and to the high unbalance between both classes (more than 90% of diabetic patients do not develop DR).

We are constructing a methodology to take advantage of the data of the new patients which are treated at the hospital. We propose to modify the set of trees that compose the FRF, which will allow updating the model without retraining the base model from the ground up.

2 Proposed method

The proposed architecture is illustrated in Fig. 1. Each time a sufficiently large set of new cases is collected, the updating method is applied to improve the classification model.

As a first dynamic component, we consider the ensemble voting procedure of the fuzzy random forest. Two methods are usually applied: majority and weighted voting. They have both been studied.

* PhD advisor: Aïda Valls Mateu

As a second dynamic component, we have the new data collected for updating the rules in the random forest. To improve results, the use of previous miss-classified examples (i.e. errors) is proposed. Three different ways of dealing with errors are studied, named: *no errors*, *errors* and *all errors*, depending on which error examples are used during the dynamic updating.

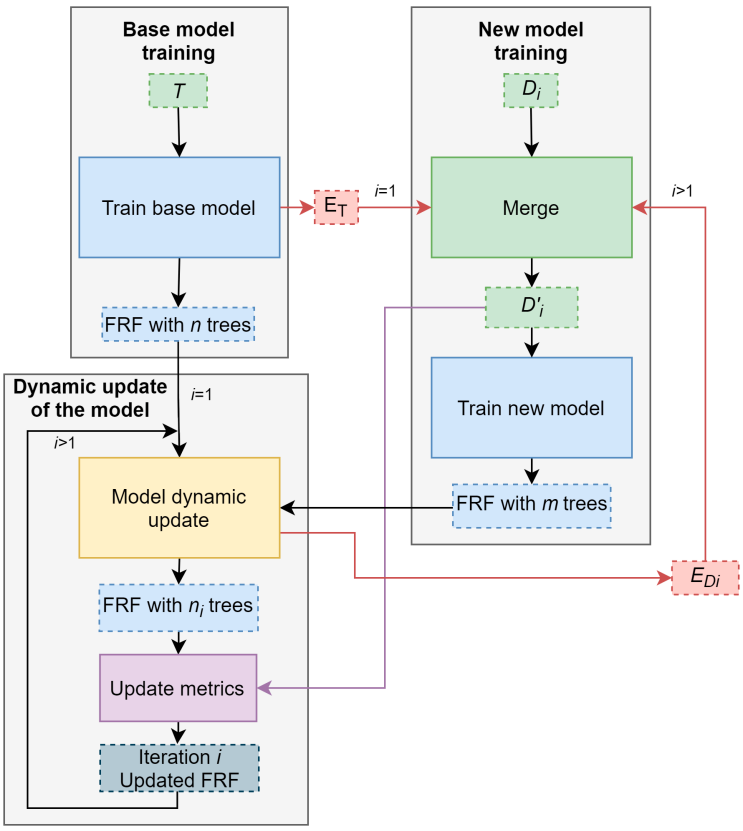


Fig. 1. Architecture of the iterative learning of Fuzzy Random Forests

The overall updating architecture is composed of three steps. The first one is not iterative, and the other two steps are run in iterations each time the model has to be updated. The three steps are briefly explained next.

1. **Base model training:** The first stage consists on training the base model with a large training dataset, T, obtaining n fuzzy decision trees, where n is a large number, usually more than 100. During the construction process, the out-of-bag samples of each fuzzy decision tree, are used to compute for each of them their specificity and sensitivity. Those metrics are used in the weighted voting, and in the update process. The training dataset can also be used for testing, and the samples that are not correctly classified are stored in E_T. Those errors samples E_T are used in the following stage.

2. **New model training:** Every time enough new samples D_i have been gathered, around 200 samples, a new training iteration i is performed. The merge process generates the dataset used to update the fuzzy random forest, D'_i, which depends on the method version. The *errors* version merges the errors data from previous iterations with D_i. For the first iteration $i = 1$, the E_T errors samples are merged. For further iterations, the E_{Di} samples generated in the third step of the method are merged. The *all errors* version also merges the training data D_i from previous iterations. Finally, the D'_i samples are used to train a new fuzzy random forest with a lower amount of trees m, around 20. Their out-of-bag samples are also used to compute the aforementioned metrics for each of the new trees.
3. **Dynamic update:** The m fuzzy decision trees trained in the previous step are used to update the current model. They are added to it, and to improve its performance, the worst fuzzy decision trees are removed. This is fixed by a certain percentage p. To sort the trees and keep the best ones, the weighted balanced accuracy is used. It is defined as an average between specificity and sensitivity with a weighting factor α.
   After pruning the worst trees, an additional update weights process can optionally be performed. The weights computed using the out-of-bag samples are updated with the training data D'_i of the current iteration.
   The resulting fuzzy random forest is set as the current model, and it is taken as the new model to be used until a new set of cases is available, and a new update iteration starts.
   The errors of the updated fuzzy random forest model on the D'_i dataset may be also retrieved and stored in E_{Di} as it was done for the base model with E_T, so they can be used in subsequent iterations.

The use of error samples has two purposes. On the one hand, to increase the size of the training set D'_i and, on the other hand, to show again this wrongly classified cases to the model in order to be able to build new rules that cover them appropriately.

3 Experimental results

Experiments are mainly done with the DR dataset obtained from the hospital, which is continuously increased with the new visit's to the patients. However, to validate the methods proposed, we have also used the occupancy dataset [1] from the UCI public repository, in which the occupancy of an office room is predicted.

The data from both problems is split in 3 different datasets: training, validation and testing. The validation set is split in different batches to simulate the new data that continuously arrives to the system. While, the testing set is used after each iteration to check the performance of the updated FRF.

Fig. 2 shows the evolution of the updated model in the test set after each iteration. The metrics on the DR dataset gradually improve during all iterations. Moreover, the sensitivity is the metric which increases the most, as it was desired. In the occupancy results, the sensitivity gradually improves, and the specificity ends slightly decreasing. Even though it is not as desired as improving both metrics, it is still desired for our use case.

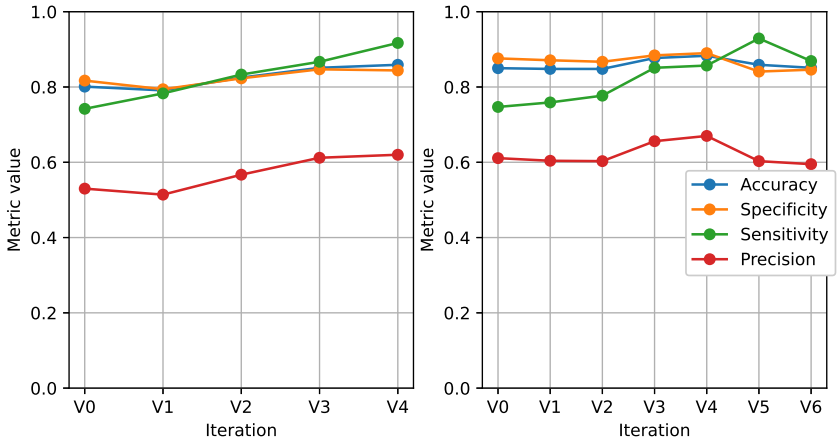


Fig. 2. Metrics on the test set. Diabetic Retinopathy (left) and Occupancy (right)

Acknowledgement. This work has been funded by the research projects PI21/00064 and PI18/00169 from Instituto de Salud Carlos III & FEDER funds. The University Rovira i Virgili also supports this work with project 2020PFR-B2-61. The author has a pre-doctoral FI grant (2021 FI_B 00139) from Generalitat de Catalunya.

References

[1] L.M. Candanedo and V. Feldheim. Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models *Energy and Buildings*, 112:(28–39), 2016.

[2] P. Romero-Aroca, A. Valls, A. Moreno, R. Sagarra-Alamo, J. Basora-Gallisa, E. Saleh, M. Baget-Bernaldiz and D. Puig. A Clinical Decision Support System for Diabetic Retinopathy Screening: Creating a Clinical Support Application. *Telemedicine and E-Health*, 25(1):31-–40, 2019.

[3] E. Saleh, A. Valls, A. Moreno, P. Romero-Aroca, V. Torra and H. Bustince. Learning Fuzzy Measures for Aggregation in Fuzzy Rule-Based Models. *Lecture Notes in Computer Science*, 11144:114-–127, 2018.

Monocular Depth Estimation with Self-supervised Graph Convolutional Network

Armin Masoumian *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili Tarragona, Spain
armin.masoumian@urv.cat

1 Abstract

Depth estimation is a challenging task of 3D reconstruction to enhance the accuracy sensing of environment awareness. Recently, convolutional neural networks (CNN) have demonstrated their extraordinary ability to estimate depth maps from monocular videos. However, traditional CNN does not support a topological structure, and they can work only on regular image regions with determined size and weights. On the other hand, graph convolutional networks (GCN) can handle the convolution on non-Euclidean data, and it can be applied to irregular image regions within a topological structure. Therefore, to preserve object geometric appearances and objects locations in the scene, in this work, we aim to exploit GCN for a self-supervised monocular depth estimation model. Our model consists of two parallel auto-encoder networks: the first is an auto-encoder which extract the feature from the input image and on multi-scale GCN to estimate the depth map. In turn, the second network will be used to estimate the ego-motion vector (i.e., 3D pose) between two consecutive frames based on ResNet-18. The estimated 3D pose and depth map will be used to construct the target image.

2 Introduction

In the Artificial Intelligence (AI) field, especially deep learning (DL) networks have accomplished high performance in various depth estimation and ego-motion prediction tasks, and nowadays, it is steeply expanding. The importance of depth estimating, as a pull factor for the entry of modern technologies into self-driving vehicles [1], object distance prediction [7]. Besides, depth maps can be used for underwater machine vision and robotic perception [9].

The stereo vision system is one of the common techniques is used for depth estimation. However, in order to save cost and computational resources, many

* PhD advisors: Prof. Domenec Puig and Dr. Hatem A. Rashwan

methods have been presented to perform depth estimation based on a monocular camera. The monocular depth estimation methods can be divided into two categories in terms of the learning approach: supervised learning methods [3] and unsupervised learning methods [4]. Most of existing DL monocular depth estimation networks use convolutional neural networks (CNN) to extract the feature information. However, CNN is limited, since it does not consider the characteristics of the geometric depth information and object location and contextual features in the scene. Besides, there is recently a need to extend deep neural models from Euclidean domains achieved by CNNs to non-Euclidean domains [2]. Thus, the research community has started to observe the importance of DL networks based on graphs [6]. The effectiveness of the graph convolution network (GCN) has been proved in processing graph data on the tasks of classification and segmentation. Thus, in this work, we propose a novel architectural DL network based on GCN, that can help to advance monocular depth estimation.

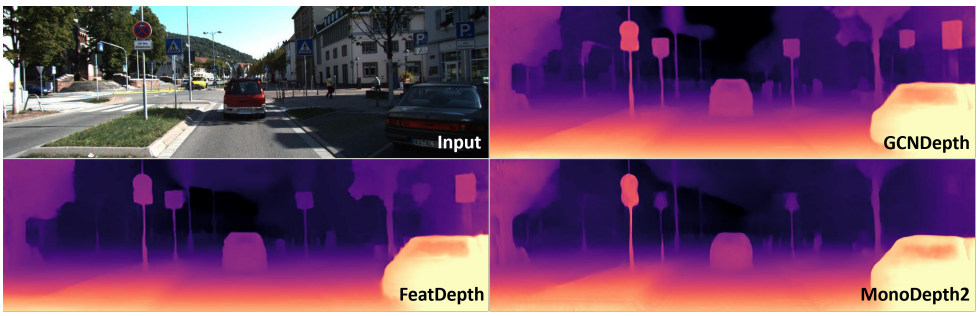


Fig. 1: Depth from a single image. GCNDepth (our self-supervised model), produces high quality depth maps with clear background and sharp edges compare to state of the art self-supervised depth estimation.

3 Summary

Based on the brief survey above, and to avoid depending on ground-truth and more generalized monocular depth estimation, we will propose a self-supervised learning approach in this work. Our method will estimate the depth images and the ego-motion to increase the constraints of depth prediction. For monocular depth estimation, the relationship between object location and visual and contextual features in the scene is significant to preserve the objects' boundaries. Most self-supervised monocular depth estimation methods [5] are based on CNN-based networks that extract appearance visual features from whole scene images. However, in most cases, CNN-based networks yield blurred edges and boundaries of the objects. We used a standard CNN

encoder for visual feature extraction and a GCN decoder for reconstructing depth maps. The reason for using GCN as a decoder network is to improve the detection of sharp boundaries and reduce the background noise to compute precise depth maps with full objects details compared to the self-supervised state-of-the-art model. For the CNN-based encoder, most monocular depth estimation used The ResNet-50 network as a backbone for feature extraction, and they achieved high performance. Thus, we similarly use ResNet-50 for the depth estimation network in our encoder. For ego-motion estimation, we used the same network proposed in [8] that is based on ResNet-18 as a backbone. In order to obtain more structural details in the scene, our approach will use a combination of different warping errors proposed in the state-of-the-art, such as the reconstruction error presented in [10] to minimize the errors in the reconstructed image, the photometric reprojection error proposed in [5] to optimize the values which provide matching pixel intensities between the target and reconstructed images. Finally, a combination between discriminative and curvature errors [8] to highlight geometric characteristics of the objects and textured regions in the scene image.

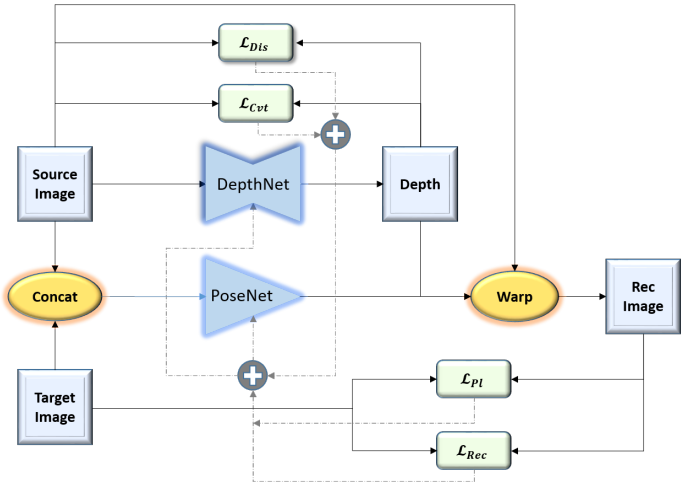


Fig. 2: Schematic illustration of the whole framework

4 Overall Pipelines

The proposed method consists of two main networks. The first network, called DepthNet. The source image is an input of the DepthNet, and the output is the depth map. The second network is PoseNet, a pose predictor to estimate the ego-motion vector of the source and the target images (in our case, a consecutive image). The output of PoseNet is the relative pose between the

source and target images. These two main networks provide geometry information to provide point-to-point correspondences of the reconstructed image. The whole architecture of our model is illustrated in Fig. 2.

5 Discussion

The performances of our model compared with the state-of-the-art solutions is summarized in Table 1. As shown in Table 1, the GCNDepth method achieved the highest performance in terms of Abs-Rel, Sq-Rel, second and third accuracy of (δ_2,δ_3) evaluation metrics. In addition, the proposed method also achieved second best results in RMSE, RMSE-Log and first accuracy of (δ_1) with a slight difference of 0.003 with RMSE-log, and 0.5% with δ_1 compared to the highest results achieved by [8]. In general, the model of Featdepth [8] and our model, GCNDepth, provided comparable results and they outperform the other tested methods.

Although, the Featdepth model achieved similar results to our model, the GCNDepth model yields a 40% reduction in the number of trainable parameters compared to the Featdepth model. Where the GCNDepth model has trainable parameters of $48,220,954$, in turn the Featdepth model has $79,681,406$. Since the Featdepth model has an extra deep feature network for feature representation learning to cope with the geometry problem of self-supervision depth estimation. The comparable results show that the use of GCN in reconstructing the depth images can improve the photometric error that appeared in the self-supervision problem without using the feature network as proposed in [8]. The results have shown in Table 1 supported that the use of GCN in estimating depth maps from a monocular video can yield depth maps outperforming or matching the state of the art on the KITTI dataset.

Table 1:
Comparison of different methods on KITTI dataset. Best results are in bold blue and second best results are in bold red color.

	Lower Better				**Higher Better**		
Method	**Abs-Rel**	**Sq-Rel**	**RMSE**	**RMSE-Log**	$\boldsymbol{\delta_1}$	$\boldsymbol{\delta_2}$	$\boldsymbol{\delta_3}$
Monodepth2 [5]	**0.115**	0.882	4.701	0.190	0.879	**0.961**	**0.982**
FeatDepth [8]	**0.104**	**0.729**	**4.481**	**0.179**	**0.893**	**0.965**	**0.984**
GCNDepth	**0.104**	**0.720**	**4.494**	**0.181**	**0.888**	**0.965**	**0.984**

Acknowledgement. This research has been possible with the support of the Secretariad Universitatsi Recercadel Departamentd Empresai Coneixement de la Generalitat de Catalunya (2020 FISDU 00405).

References

[1] Badue, Claudine and Guidolini, Rânik and Carneiro, Raphael V and Azevedo, Pedro and Cardoso, Vinicius B and Jesus, Luan F R and Berriel, Rodrigo F and Paixão, Thiago M and Mutz, Filipe and Oliveira-santos, Thiago and Souza, Alberto F De Self - Driving Cars A Survey, 2017

[2] Bronstein, Michael M and Bruna, Joan and Lecun, Yann and Szlam, Arthur and Vandergheynst Geometric deep learning: going beyond Euclidean data *arXiv:1611.08097v2, 1–22, 2017.*

[3] Eigen, David and Puhrsch, Christian and Fergus, Rob Depth Map Prediction from a Single Image using a Multi-Scale Deep Network *Advances in Neural Information Processing Systems, 2366–2374, 2014.*

[4] Garg, Ravi and Vijay Kumar, B. G. and Carneiro, Gustavo and Reid, Ian Unsupervised CNN for single view depth estimation Geometry to the rescue *Lecture Notes in Computer Science, 9912, 2016.*

[5] Godard, Clément and Mac Aodha, Oisin and Firman, Michael and Brostow, Gabriel Digging Into Self-Supervised Monocular Depth Estimation *arXiv preprint arXiv:1806.01260, 2018.*

[6] Kipf, Thomas N and Welling, Max Semi-supervised classification with graph convolutional networks *arXiv preprint arXiv:1609.02907, 2016.*

[7] Masoumian, Armin and Marei, DG and Abdulwahab, Saddam and Cristiano, Julian and Puig, Domenec and Rashwan, Hatem A Absolute distance prediction based on deep learning object detection and monocular depth estimation models *CCIA2021, 339-325, 2021, IOS Press.*

[8] Shu, Chang and Yu, Kun and Duan, Zhixiang and Yang, Kuiyuan Feature-metric loss for self-supervised learning of depth and egomotion *European Conference on Computer Vision, 572–588, 2020, Springer.*

[9] Ye, Xinchen and Li, Zheng and Sun, Baoli and Wang, Zhihui and Xu, Rui and Li, Deep joint depth estimation and color correction from monocular underwater images based on unsupervised adaptation networks *IEEE Transactions on Circuits and Systems for Video Technology, 30-11, 3995–4008, 2019.*

[10] Zhao, Hang and Gallo, Orazio and Frosio, Iuri and Kautz, Jan Loss functions for image restoration with neural networks *IEEE Transactions on computational imaging, 3, 47–57, 2016, IEEE.*

Contributions to GDPR compliance by means of Smart Contracts

Cristòfol Daudén-Esmel *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
cristofol.dauden@urv.cat

1 Introduction

The rapid advancement and development of new digital technologies has changed the dynamics of our daily lives by providing us with new services and products. Among the services, we can stress the social connectivity, information storage and location (GSP and mapping). Regarding the products, Smartwatches and Smart Home Devices with an Intelligent Personal Assistant (IPA) are two examples that are becoming more popular worldwide. These services and devices generate huge amounts of information, which is processed by the Service Providers (SPs) in order to improve and develop new products. We cannot however ignore that also represents an important source of revenue for the SPs. As a result, their products can be cheaper or even free [1]

The processing of the aforementioned information may result in extraction of sensitive information which can jeopardise the users' privacy. Thus, the EU took a decisive step with the General Data Protection Regulation (GDPR) [2], that came into effect from May 2018, in order to protect users' rights. The GDPR wants to mitigate the abuse of massive collection and processing of users' personal data. The regulation guarantees specific privacy rights to Data Subjects (physical or legal entities to which the personal data belongs) ensuring that personal data "can only be gathered legally, under strict conditions, for a legitimate purpose"; as well as bringing full control back to the data owners.

Under GDPR, companies are required to prove compliance in case of suspicion of a violation or when a Data Subject (DS) lodges a complaint with the Supervisory Authority (SA). However, the legislative text does not specify how to transparently demonstrate that the information collected and its processing fulfills with the regulation. In the same way, DSs need tools to know and control what happens with their personal data. So, individuals have no tools to know transparently and easily which data is being collected and pro-

* PhD advisors: Jordi Castellà-Roca and Alexandre Viejo

cessed and for which purposes. As a result, DSs are mostly limited to giving their consent beforehand, in a way that is based on an abstract clause. In this regard, the current GDPR-compliance verification architectures generally depends on each service provider, i.e. they are specific and centralized for each of them. Due to this reason, critical concerns on the lack of transparency have been imposed [3].

It is therefore necessary to deploy a framework in order to enable the agreement verification between the users (DS), Data Controllers (DC) and Data Processors (DP) in relation to the data custody and processing. At the same time, the users should be capable to know and control which data is being collected, who is processing it and for which purposes. From the DS point of view, the main benefit is a way to manage his personal data, which does not depend on the DC, i.e. the tool can be used to manage all agreements with SPs. In addition, from a DC and DP point of view, the main benefit is a proof that can be presented to SAs showing that data was obtained and processed in a GDPR compliant way. So, the proof should have the following properties: i) public access; ii) verifiable; iii) authentic; iv) immutable and; v) non-repudiable. According to these properties, some authors have proposed the use of Smart Contracts (SCs) implemented over the blockchain technology (BC) as a general-purpose data management [4–10]. This is a promising technology in GDPR-compliant personal data management.

2 Contributions

Our first contribution to this topic is a lightweight blockchain-based GDPR-compliant personal data management system, which provides public access immutable evidences showing the agreements between a Data Subject and a Service Provider about DS's personal data. Compared to other existing research works, our work proposes a new conceptual design and system architecture for human-centric personal data management, by using BC and SCs technology that is in compliance with the GDPR (see Figure 1). Our work differentiates between the data collection and data processing concepts by identifying the Controller and Processor actors and treating them in a related but separate way. We also try to reduce the overhead on DSs, as if they need to have a wide knowledge on BC technology or they have to be constantly operating over our platform, it will be hard to be accepted and used by the community.

On the current work we are extending the preliminary scheme presented in [11]. In particular, the new scheme has been partially re-designed to be deployed more conveniently in a realistic setting. This includes: i) a modification in the process flow that makes the DS the main responsible of her own personal data and the initiator of the whole proposed protocol; ii) the use of the well-known XACML framework to improve the robustness of the

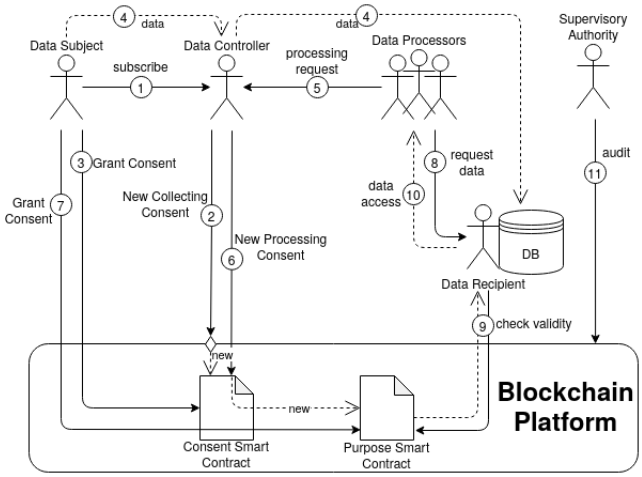


Fig. 1. System Architecture

access control process to DSs collected data; and iii) a refinement in the use of the SCs that allows us to include all the purposes of a certain DP in a single contract, thus, improving the general efficiency of the proposed system; Moreover, in the new proposal, we have done a more detailed experimental study, including the implementation of a realistic use case.

Future Work

Every actor needs an asymmetric key to use the proposed system, as digital signature is used to interact with Smart Contracts. The public key (PK) can be seen as an ID of the actor itself, so in order to keep DSs' anonymity to possible linkage attacks, a new key pair is used for every consent with a different Data Controller.
In order to make the asymmetric keys management abstract for DSs, as future work, we pretend to complement our work with a tool that allows them to generate, store and manage all asymmetric keys used to interact with the proposed system, in a transparent way. This tool must be multi-platform, secure and tramper-resistant as holds all IDs (PKs) a Data Subject uses in our system and their associated secret keys to interact with the generated agreements.

Acknowledgement. This research was supported by the European Union Regional Development Fund within the framework of the ERDF Operational Program of Catalonia 2014-2020 with a grant of 50% of the total cost eligible, under the FEM-IOT project [001-P-001682]; the European Commission (project H2020-871042 "SoBigData++"); and the Spanish Government (project RTI2018-095094-B-C21 "Consent"). The author is also supported by the Spanish Government under an FPU grant (ref. FPU20/03254).

References

[1] Asunción Esteve. The business of personal data: Google, Facebook, and privacy issues in the EU and the USA. *International Data Privacy Law*, 7(1):36–47, 03 2017.

[2] Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation) (text with eea relevance). *Official Journal of the European Union L 119*, 59:1–88, 5 2016.

[3] Shakila Bu-Pasha, Anette Katariina Alen-Savikko, Jenna-Sofia Mäkinen, Robert Guinness, and Päivi Hannele Korpisaari. Eu law perspectives on location data privacy in smartphones and informed consent for transparency. *European Data Protection Law Review*, 2(3/2016):312–323, September 2016.

[4] Laure A Linn and Martha B Koo. Blockchain for health data and its potential use in health it and health care related research. In *ONC/NIST Use of Blockchain for Healthcare and Research Workshop. Gaithersburg, Maryland, United States: ONC/NIST*, pages 1–10, 2016.

[5] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman. Medrec: Using blockchain for medical data access and permission management. In *2016 2nd International Conference on Open and Big Data (OBD)*, pages 25–30, Aug 2016.

[6] Ricardo Neisse, Gary Steri, and Igor Nai Fovino. A blockchain-based approach for data accountability and provenance tracking. In *ARES '17: Proceedings of the 12th International Conference on Availability, Reliability and Security*, pages 1–10, August 2017.

[7] S. Wang, Y. Zhang, and Y. Zhang. A blockchain-based framework for data sharing with fine-grained access control in decentralized storage systems. *IEEE Access*, 6:38437–38450, 2018.

[8] Christian Wirth and Michael Kolain. Privacy by blockchain design: A blockchain-enabled gdpr-compliant approach for handling personal data. In *Reports of the European Society for Socially Embedded Technologies (EUSSET)*, 2018.

[9] Benedict Faber, Georg Michelet, Niklas Weidmann, Raghava Rao Mukkamala, and Ravi Vatrapu. Bpdims:a blockchain-based personal data and identity management system. In *Conference: Hawaii International Conference on System Sciences*, January 2019.

[10] N. B. Truong, K. Sun, G. M. Lee, and Y. Guo. Gdpr-compliant personal data management: A blockchain-based solution. *IEEE Transactions on Information Forensics and Security*, 15:1746–1761, 2020.

[11] Cristòfol Daudén-Esmel, Jordi Castellà-Roca, Alexandre Viejo, and Josep Domingo-Ferrer. Lightweight blockchain-based platform for gdpr-compliant personal data management. In *5th IEEE International Conference on Cryptography, Security and Privacy, CSP 2021, Zhuhai, China, January 8-10, 2021*, pages 68–73, 2021.

Deep learning-Based Approach for Retinal Lesions Segmentation in Eye Fundus Images

Moahammed Yousef Salem Ali *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
mohammedyousefsalem.ali@estudiants.urv.cat

1 Introduction

Early diagnosis of retinal lesions helps reduce the risk of visual loss and blindness. Ophthalmologists inspect eye fundus images to detect the signs of common eye diseases like diabetic retinopathy (DR) and glaucoma. Figure 1 shows the most common types of lesions that may affect the retina. The yellow spots in the retina region stand for hard exudates (HX), pale yellow or white areas with ill-defined edges stands for soft exudates (SX), tiny outpouchings of blood stands for microaneurysms, while the bleeding that occurs in the retina is known as haemorrhages.

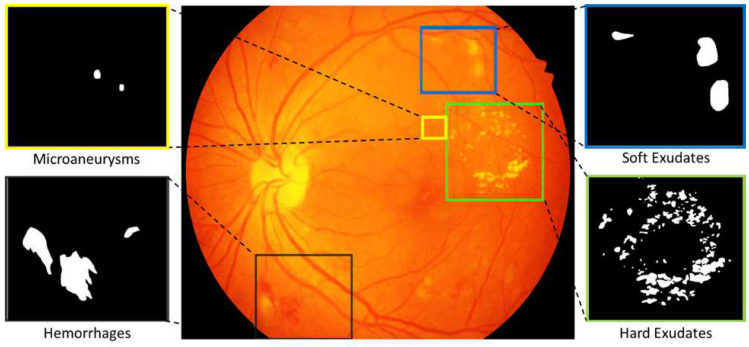


Fig. 1. Retinal lesions types.

Indeed, ophthalmologists dedicate many hours to perform manual analysis of hundreds of fundus images, which represents a high cost considering manpower needed and salaries [1], [5]. On the other hand, artificial intelligence-based computer-aided diagnosis (CAD) systems, if trained properly, can an-

* PhD advisor: Dra. Aida Valls, Dr. Mohamed AbdelNasser, and Dr. Marc Baget

alyze hundreds of fundus images and provide a diagnosis as experienced ophthalmologists [2].

In this research, we leverage emerging deep learning technologies such as U-Net and Fully Convolutional Network (FCN) to automatically detect and segment various lesions in eye fundus images.

2 Methodology

In this study, we use a deep learning-based model termed gated skip connections [4] to distinguish and segment hard and soft exudates properly in fundus images. The model comprises five encoder blocks with two convolutional layers and five decoder blocks with four convolutional layers (Figure 2). In this model, an efficient skip connections technique is combined with the U-Net architecture's decoder to retrieve eye-lesion-relevant information while disregarding irrelevant features.

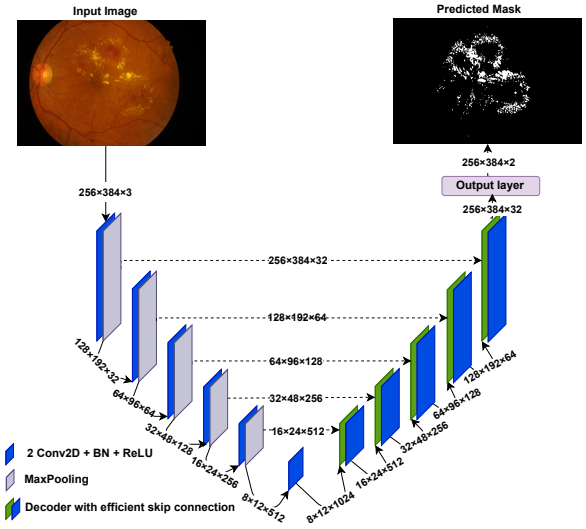


Fig. 2. Eye lesion segmentation using deep gated skip connections.

The Indian Diabetic Retinopathy Image Dataset (IDRiD) [3] is used to train and evaluate the eye lesion segmentation algorithm. IDRiD provides 81 fundus images of the retina with excellent annotations for the optic disk, microaneurysms, hemorrhages, hard and soft exudates. The fundus photos are 4288 x 2848 pixels in size. The database is divided into two standard sections for training and testing: 54 photos for training and 27 images for testing. It should be mentioned that the photos may depict a variety of different sorts of lesions. To take benefit from the high resolution of the images while minimizing

computing costs, we split each image into 12 splits during the training and testing phases. Data augmentation techniques are used to increase the number of training images. The model is trained using the binary cross-entropy loss function and the ADAM optimizer.

Two segmentation models for eye lesions have been trained: one of them for hard exudates and another for soft exudates.

3 Preliminary results and future work

Table 1 presents the results of the two segmentation models and the state-of-the-art models. With a hard exudates segmentation task (HX), the proposed eye lesion segmentation model obtains F1-score and area under the precision-recall curve (AUPR) of 75.9 and 84.8%, respectively. The F1-score of the proposed model is 0.7 points better than the method proposed by Xiao et al. [6]. A soft exudates segmentation task (SX) achieves an F1-score and AUPR of 68.7 and 75.0%, respectively. As one can see, the proposed model outperforms HEDNet+cGAN [6], and Saha et al. [7].

Table 1. Preliminary results and comparison

Lesion	**Method**	**Metrics (%)**	
		F1	**AUPR**
Hard exudates	HEDNet+cGAN [6]	69.0	84.1
	Saha et al. [7]	-	**87.0**
	Proposed	**75.9**	84.8
Soft exudates	HEDNet+cGAN	44.0	48.4
	Saha et al.	-	71.0
	Proposed	**68.7**	**75.0**

Figure 3 shows a sample of results of HX (top) and SX (bottom). The segmentation models views result is much closer to the ground truth.

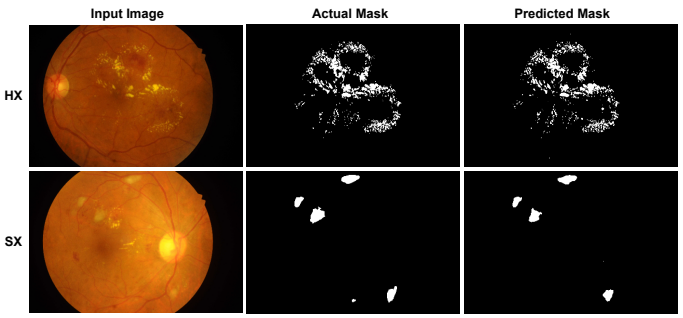


Fig. 3. Segmentation results.

The future work will include using meta-learning deep-learning-based techniques to improve the segmentation results and performing segmentation of the other types of eye lesions like microaneurisms and haemorrhages.

Acknowledgement. This work has been funded by the research project PI18/00169 from Instituto de Salud Carlos III & FEDER funds. The University Rovira i Virgili also supports this work with project 2019PFR-B2-61.

References

[1] Mary, Viola Stella, Elijah Blessing Rajsingh, and Ganesh R. Naik. "Retinal fundus image analysis for diagnosis of glaucoma: a comprehensive survey." IEEE Access (2016).

[2] Jani, Kuntesh, Rajeev Srivastava, Subodh Srivastava, and Animesh Anand. "Computer aided medical image analysis for capsule endoscopy using conventional machine learning and deep learning In 2019 7th International Conference on Smart Computing Communications (ICSCC), pp. 1-5. IEEE, 2019.

[3] Porwal, Prasanna, Samiksha Pachade, Ravi Kamble, Manesh Kokare, Girish Deshmukh, Vivek Sahasrabuddhe, and Fabrice Meriaudeau. "Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research." *Data 3, no. 3 (2018): 25.*

[4] Jabreel, M and Abdel-Nasser, M " Promising crack segmentation method based on gated skip connection" *Electronics Letters 56, no. 10 (2020): 493-495.*

[5] ALI, Mohammed Yousef Salem, Mohamed ABDEL-NASSER, Mohammed JABREEL, Aida VALLS, and Marc BAGET. "Segmenting the Optic Disc Using a Deep Learning Ensemble Model Based on OWA Operators." *ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT (2021): 305.*

[6] Xiao, Qiqi, Jiaxu Zou, Muqiao Yang, Alex Gaudio, Kris Kitani, Asim Smailagic, Pedro Costa, and Min Xu. "Improving Lesion Segmentation for Diabetic Retinopathy using Adversarial Learning." *In International Conference on Image Analysis and Recognition*, pp. 333-344. Springer, Cham, 2019.

[7] Saha, Oindrila, Rachana Sathish, and Debdoot Sheet. "Learning with multitask adversaries using weakly labelled data for semantic segmentation in retinal images." *In International Conference on Medical Imaging with Deep Learning*, pp. 414-426. PMLR, 2019.

Fundus Image Quality Assessment Based on Deep Autoencoder Networks

Saif khalid *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili Tarragona, Spain
saifkhalidmusluh.al_khalidy@estudiants.urv.cat

1 Abstract

Fundus image quality is critical to diagnosing retinal diseases since image clarity is significant in classifying such images. This work presents a new field-friendly multitasking framework for automatically interpreting base image quality based on the autoencoder network used to reconstruct the input image. The proposed system provides an interpretable quality assessment and quality visualization. In particular, the present application can detect the optical disc and pure structures as features to help the evaluation by coding. The experimental results have shown the superiority of the proposed approach over various modern methods.

2 Introduction

Modern medicine highlights big data to assess fundus image quality based on the human visual system. In ophthalmology, the use of fundus photography has been highlighted, which has given rise to indispensable applications of portable fundus cameras. However, in fundus photography, image quality is more susceptible to general quality distortions, such as color distortion, uneven lighting, low contrast, and stuttering. Digital fundus imaging is used to diagnose various eye disorders such as diabetic retinopathy (DR) [1], cataract [2], age-related macular degeneration (AMD) [3], and glaucoma [4].

Scientists focus on ways to obtain effective medical help for a large number of patients. However, the number of eye specialists available needed fails to meet the current demand . To address the lack of the required ophthalmologists, telemedicine [5], and computer-aided diagnosis (CAD) [6] can be used at eye diseases diagnosis and prognosis.

All CAD systems of eye disease diagnostic systems are based on the quality of retinal images. The results of CAD systems with low-resolution images

* PhD advisor: Hatem A. Rashwan and and Mohamed Abdel-Nasser Domenec Puig

degrade their decision-making performance. Thus, a trustworthy assessment of retinal fundus image quality is needed to improve the early detection of eye diseases. In this work, we propose a framework based on deep learning techniques, mainly a deep autoencoder network, to develop a reliable fundus image assessment. The model consists of two cascading networks: an autoencoder network for self-supervision based on image reconstruction and a deep CNN classifier for classifying the quality of the input image. In the autoencoder network, a multi-layer encoder will be used to extract local and global features related to the quality of retinal images and decode them to reconstruct the same input image. Then the features obtained from the encoding network are fed to the classifier to classify the quality of the network input images.

In most CAD diagnosis systems After training the model, we analyze their representations via attribution and other interpretability methods. Our contributions to this paper are as follows:

- We propose an auto-encoding network to correctly recognize the representative depth features of fundus images via the cryptographic network. The decoder part is used to reconstruct the input bottom image.
- We suggest a CNN classifier fueled by the features learned by the encoder network to classify input fundus images as gradable or ungradable.
- We suggest using a measure of mean square error (MSE) as a loss function To train the automatic encryption network. The MSE loss function calculates the sum of Square the distance between the input image and the image reconstructed by the decoder. Also, we use the binary entropy loss function to train the CNN classifier.
- We propose to integrate the losses of the two autoencoder networks and the CNN classifier into a single learning framework to solve the fundus image gradability problem.
- We apply feature attribution and other interpretability methods to understand the representation of the fundus images in both models.
- Our interpretability analysis indicates that the autoencoder loss helps the classifier focus more on the relevant structures of the fundus images, such as the fovea, optic disc, and main blood vessels. The normal model, on the other hand, uses more arbitrary input regions to determine the gradability of the image.

3 Proposed Model

Figure 1 gives a high level overview of the training and testing phases of our proposed model. In the first training model,we used an autoencoder net- work consisting of two serial networks: the encoder and the decoder.We used the encoder network to extract the high-level features of the model input fundus

images. Next stage after feature extraction these features will be fed to the decoder network so as to rebuild the same input image again. In the next stage,another network, the classifier network, will be fed with the features obtained from the autoencoder network to classify the retinal image quality into two categories: gradable and ungradable. The size of the input image was resampled to 480x480. In the testing phase of the model, we used only the trained encoder and classifier grid in order to classify the image quality of the fundus mesh in addi- tion to entering it into the interpretation phase, which is an important phase in testing medical images and classifying them as gradable and ungradable .

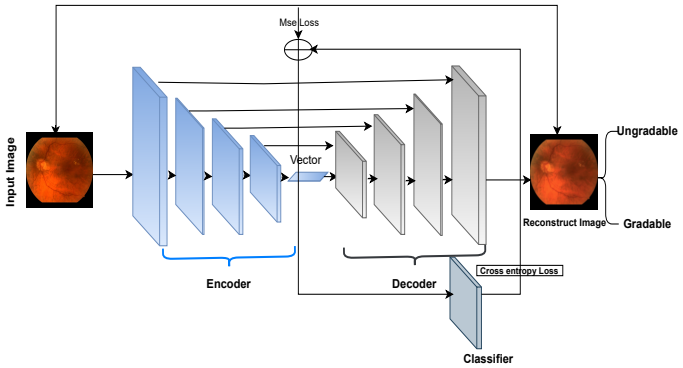


Fig. 1. General overview of the autoencoder model in train stage.

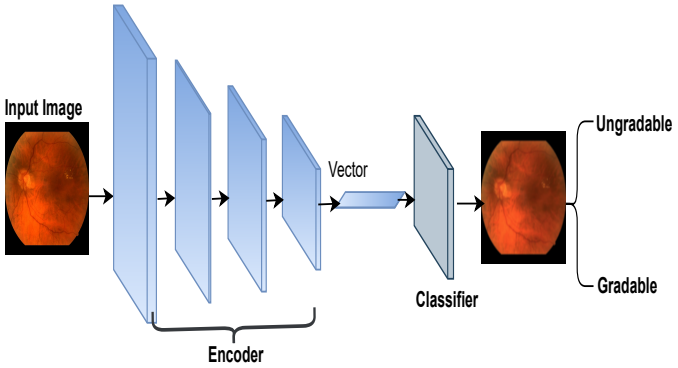


Fig. 2. General overview of the autoencoder model in test stage.

4 results

Based on the two datasets, Table 1 and 2 show the results (i.e, Accuracy, Sensitivity, Specificity, Precision and $F1$ score) of MCF-Net and four variations of the proposed systems with the four loss functions. As shown in Table 1, the proposed model with its four variations outperformed the performance of MCF-Net in terms of the five evaluation matrices. Among them, our model with MSE as a loss function yielded the best results with $F1$ score, sensitivity and specificity of 0.88, 0.83 and 0.91, respectively. For instance, our model with MSE yielded an improvement of 8% with $F1$ score compared to the MCF-Net. In turn as shown in Table 2 and with the second dataset EyeQ, the proposed model and its variations also outperformed the results with MCF-Net. Our model with MSE achieved significant improvements of 16%, 10% and 38% with $F1$ score, precision and specificity, respectively. Besides, a small improvement of around 1% with sensitivity.

Table 1. Comparison between the proposed model and MCF-Net [7] on the Eypces dataset [8]

	Accuracy	**Sensitivity**	**Specificity**	**Precision**	**F1-Score**
MCF-Net Model	0.81	0.64	**0.95**	0.84	0.80
Our Model - SSIM Loss	0.815	**0.95**	0.65	0.84	0.82
Our Model - MS-SSIM Loss	0.86	0.94	0.76	0.87	0.86
Our Model - MAE Loss	0.85	0.84	0.86	0.85	0.85
Our Model - MSE Loss	**0.875**	0.83	0.91	**0.88**	**0.88**

Table 2. Comparisons of the proposed model and state-of-the-arts on (EyeQ) dataset [7]

	Accuracy	**Sensitivity**	**Specificity**	**Precision**	**F1-Score**
MCF-Net Model	0.865	0.946	0.51	0.80	0.75
Our Model - SSIM Loss	0.93	0.94	0.90	0.88	0.90
Our Model - MS-SSIM Loss	0.935	0.93	**0.91**	**0.94**	**0.93**
Our Model - MAE Loss	0.94	0.95	0.88	0.90	0.91
Our Model - MSE Loss	**0.942**	**0.954**	0.89	0.90	0.91

5 Interpretation of Model Features

the proposed model Shown in the Figure 3 interprets fundus images with scores with explanability to help doctors and medical care workers distinguish gradable and non-estimable images based on grades and interpretations

that have been adopted after sending them to the General Hospital in Tarragona,spain and presenting them to a group of experts to confirm the results of the model classification Which has proven to be successful and superior to the normal model.

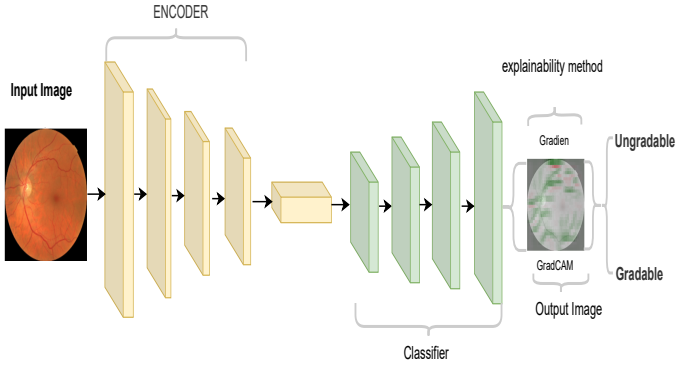


Fig. 3. General overview of the autoencoder model with explanation.

We use various interpretability methods to understand the Normal and Autoencoder models and compare their internal representations. Our approach employs:

- Saliency map methods such as GradCAM visualizations to understand the relevance of the input regions [9] .

6 Conclusions and future work

In our research paper, we proposed a supervised deep learning model based on an autoencoder network. The autoencoder network is able to generate the same network input as fundus images to correctly identify the visual features of eye image quality. Our model also includes a classifier fed by features extracted from the encoder network to rank the quality from the retinal image to Gradable and Ungradable. In addition, by analyzing the interpretability analysis, we show that the gradability models mainly focus on the presence and type of blood vessels in the fundus image. Other key structures such as the optic disk and macula seem to play a lesser role than expected. Finally, via this analysis, we also found that the addition of the decoder and corresponding loss helps the proposed model focus more on relevant structures of the fundus image

References

[1] M. R. K. Mookiah, U. R. Acharya, C. K. Chua, C. M. Lim, E. Ng, and A. Laude, "Computer-aided diagnosis of diabetic retinopathy: A review," *Computers in biology and medicine*, vol. 43, no. 12, pp. 2136–2155, 2013.

[2] L. Guo, J.-J. Yang, L. Peng, J. Li, and Q. Liang, "A computer-aided healthcare system for cataract classification and grading based on fundus image analysis," *Computers in Industry*, vol. 69, pp. 72–80, 2015.

[3] M. U. Akram, A. Tariq, S. A. Khan, and M. Y. Javed, "Automated detection of exudates and macula for grading of diabetic macular edema," *Computer methods and programs in biomedicine*, vol. 114, no. 2, pp. 141–152, 2014.

[4] G. D. Joshi, J. Sivaswamy, and S. Krishnadas, "Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment," *IEEE transactions on medical imaging*, vol. 30, no. 6, pp. 1192–1205, 2011.

[5] L. Shi, H. Wu, J. Dong, K. Jiang, X. Lu, and J. Shi, "Telemedicine for detecting diabetic retinopathy: a systematic review and meta-analysis," *British Journal of Ophthalmology*, vol. 99, no. 6, pp. 823–831, 2015.

[6] C. Sinthanayothin, J. F. Boyce, T. H. Williamson, H. L. Cook, E. Mensah, S. Lal, and D. Usher, "Automated detection of diabetic retinopathy on digital fundus images," *Diabetic medicine*, vol. 19, no. 2, pp. 105–112, 2002.

[7] H. Fu, B. Wang, J. Shen, S. Cui, Y. Xu, J. Liu, and L. Shao, "Evaluation of retinal image quality assessment networks in different color-spaces," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 48–56, Springer, 2019.

[8] B. Graham, "Kaggle diabetic retinopathy detection competition report," *University of Warwick*, 2015.

[9] G. Ras, M. van Gerven, and P. Haselager, "Explanation methods in deep learning: Users, values, concerns and challenges," in *Explainable and interpretable models in computer vision and machine learning*, pp. 19–36, Springer, 2018.

Radiomics-based computer-aided diagnosis system for prostate cancer classification in MRI images

Eddardaa Ben Loussaief *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili Tarragona, Spain
Eddardaa.benloussaief@urv.cat

1 Abstract

The use of magnetic resonance imaging (MRI) in prostate segmentation, diagnosis, and treatment is critical. Using MRI images of all modalities, computer-aided diagnostic (CAD) systems based on machine learning can assist doctors in detecting prostate cancer and its aggressiveness at an earlier stage. One of the most important stages of CAD systems is automatic prostate gland delineation. With medical images, deep learning has lately exhibited encouraging segmentation outcomes. We examine the current state-of-the-art of deep learning-based techniques for prostate segmentation in MRI images and explain their benefits and shortcomings in this paper. In addition, we present a new approach for classifying prostate biopsy malignancies in MRI images. We want to leverage the segmentation results to extract deep radiomics features from MRI prostate images in this way.

2 Introduction

The most prevalent malignant tumor in men worldwide is prostate cancer. In the diagnosis and treatment of prostate cancer, accurate detection of the prostate gland utilizing medical scans is critical. Deep learning-based algorithms have made significant success in a variety of domains, including computer vision, natural language processing, and medical imaging diagnosis, according to early attempts. The potentials of deep learning-based approaches for medical imaging segmentation are still being investigated in the literature. And, as it is observed, the findings of automated prostate detection are still difficult to come by.

The main goal of this study is to compare current deep learning-based approaches for prostate cancer detection in MRI scans. Each segmentation

* PhD advisor: Dr. Mohamed Abdel-Nasser, and Dr. Domenec Puig.

model's advantages and disadvantages are highlighted. The results of segmentation on three public datasets will be presented: Promise12, ISBI Challenge2013, and ProstateX. To evaluate the performance of the prostate cancer detection algorithms, we employed the dice coefficient and Hausdorff Distance evaluation measures. The use of deep radiomics features acquired from MRI images to distinguish benign from malignant prostate cancers is our novel contribution.

3 Methodology

The recommended methodology [1]for our investigation is depicted in Figure 1 as an overview. We have trained the deep learning models to segment prostate cancer from MRI scans. The training is accomplished on a set of public datasets such as Promise12 [5], ISBI2013 [7], and ProstateX [2]. Both 2D and 3D segmentation models are essential to our strategy. We employed the U-Net [9] and 2D-Unet [3] models for two-dimensional segmentation. 3DFCN [8], 3D-Unet [4], and MS-Net [6] have all been trained for 3D segmentation.

We separated the data in our experiments into training and testing data. To boost the amount of training data, we used a data augmentation approach. To evaluate the models' performance, we compute the Dice coefficient and Hausdorff distance. The models discussed above were tested on the ProstateX citer31 dataset. This stage lays the groundwork for our ultimate goal, which is to extract deep learning-based radiomics to categorize malignant malignancies.

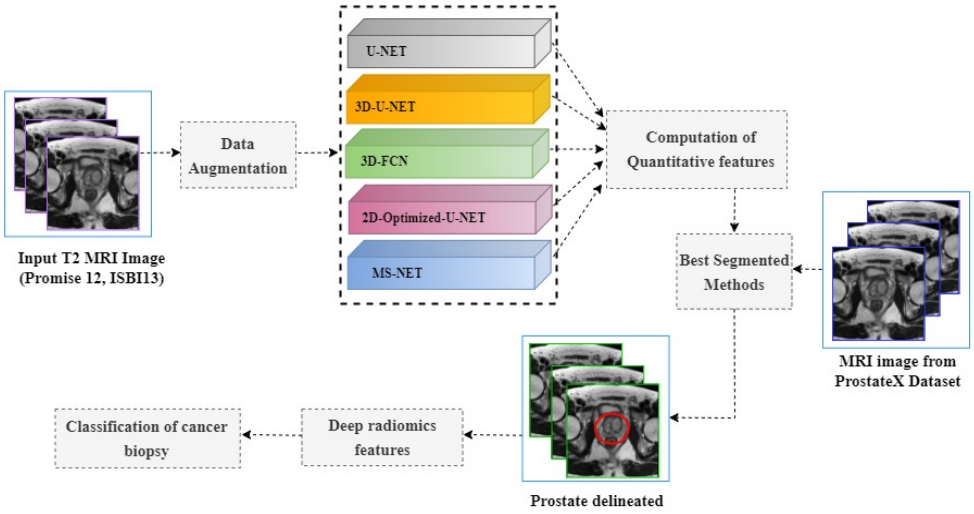


Fig. 1. The schematic illustration of the proposed methodology.

4 Results

We used the Promise12 and ISBI2013 datasets to train multiple segmentation models in order to assess their performance. The segmentation findings in terms of the Dice coefficient and Hausdorff Distance are presented in Tables 1 and 2.

Table 1. Comparing the performance of the segmentation models using ISBI2013 dataset.

Model	DSC±std	HD(mm)±std
MS-Net [6]	0.899 ±1.960	9.511 ±4.011
2D-Unet [3]	0.901 ±0.015	6.030 ±3.082
3D-Unet[4]	0.722 ±0.020	17.761 ±2.924

Table 2. Comparing the performance of the segmentation models using Promise12 dataset.

Model	DSC±std	HD(mm)±std
U-Net[9]	0.880±0.041	17.690 ±2.087
3D-FCN[8]	0.790 ±0.050	12.910 ±4.005
2D-Unet [3]	0.899 ±0.021	7.661 ±3.924

With the ISBI and Promise12 datasets, the 2D-optimised Unet delivers the best dice coefficient, as shown in Tables 1 and 2. The Hausdorff Distance (HD) of the 2D-Unet model is 6.03 mm. These findings show that the 2D-Unet model also produces accurate segmentation results on the ProsateX Dataset, as shown in Table 3.

Table 3. Segmentation result on prostatex dataset.

Model	DSC±std	HD(mm)±std
U-Net[9]	0.791±0.151	17.020 ±2.884
3D-UNet [4]	0.701 ±0.078	18.001 ±3.108
3D-FCN [8]	0.721 ±0.047	13.411 ±5.264
2D-UNet [3]	0.898±0.051	7.690 ±2.987

5 Conclusion

A comparison study of the state-of-the-art of deep learning-based segmentation algorithms for prostate cancer in MRI images has been reported. Different measures, such as the dice coefficient and Hausdorff Distance, were utilized

to evaluate the performance of the evaluated models. The deep radiomics will next be extracted and fed into a classifier to distinguish between prostate cancer groups (e.g., benign or malignant).

References

[1] Eddardaa Ben Loussaief, Mohamed Abdel-Nasser, Domenec Puig, "Prostate cancer delineation in MRI imagesbased on deep learning: quantitativecomparison and promising perspective",Artificial Intelligence Research and Development. DOI: 10.3233/FAIA210148

[2] Geert Litjens, Oscar Debats, Jelle Barentsz, Nico Karssemeijer, and Henkjan Huisman. "ProstateX Challenge data", The Cancer Imaging Archive (2017). DOI: 10.7937/K9TCIA.2017.MURS5CL.

[3] Gillespie D, Kendrick C, Boon I, Boon C, Rattay T, Yap MH. Deep learning in magnetic resonance prostate segmentation: A review and a new perspective. arXiv preprint arXiv:2011.07795. 2020 Nov 16.

[4] Isensee F, Petersen J, Klein A, Zimmerer D, Jaeger PF, Kohl S, Wasserthal J, Koehler G, Norajitra T, Wirkert S, Maier-Hein KH. nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486. 2018 Sep 27.

[5] Litjens G, Toth R, van de Ven W, Hoeks C, Kerkstra S, van Ginneken B, Vincent G, Guillard G, Birbeck N, Zhang J, Strand R. Evaluation of prostate segmeZhu Q, Du B, Turkbey B, Choyke PL, Yan P. Deeply-supervised CNN for prostate segmentation. In2017 international joint conference on neural networks (IJCNN) 2017 May 14 (pp. 178-184). IEEE.ntation algorithms for MRI: the PROMISE12 challenge. Medical image analysis. 2014 Feb 1;18(2):359-73.

[6] Liu Q, Dou Q, Yu L, Heng PA. MS-net: Multi-site network for improving prostate segmentation with heterogeneous MRI data. IEEE transactions on medical imaging. 2020 Feb 17;39(9):2713-24.

[7] Nicholas Bloch AM. NCI-ISBI 2013 Challenge: Automated Segmentation of Prostate Structures [Internet]. The Cancer Imaging Archive; 2015. Available from: https://wiki.cancerimagingarchive.net/x/B4NEAQ

[8] Milletari F, Navab N, Ahmadi SA. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In2016 fourth international conference on 3D vision (3DV) 2016 Oct 25 (pp. 565-571). IEEE.

[9] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. InInternational Conference on Medical image computing and computer-assisted intervention 2015 Oct 5 (pp. 234-241). Springer, Cham.

Breast Tumor Segmentation in Ultrasound Image using Deep Learning Techniques

Nadeem Issam Zaidkilani *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili Tarragona, Spain
nadimkilani@gmail.com

1 Introduction

Breast cancer is one of the leading causes of cancer related death for women worldwide and poses a growing health problem, the most urgent is to diagnose breast cancer in early stage. In the last decades, computer-aided diagnosis (CAD) systems have been introduced to help for physicians. It doesn't only create and analyze images, but also becomes an assistant and helps doctors with their interpretation. Deep learning methods, especially convolutional neural networks (CNNs), have been successfully applied to lesion segmentation in breast ultrasound (BUS) images. In our research, we employ state-of-the-art deep learning-based semantic segmentation for breast tumor segmentation in ultrasound images. An example of the benign and malignant tumors is shown in Fig.1.

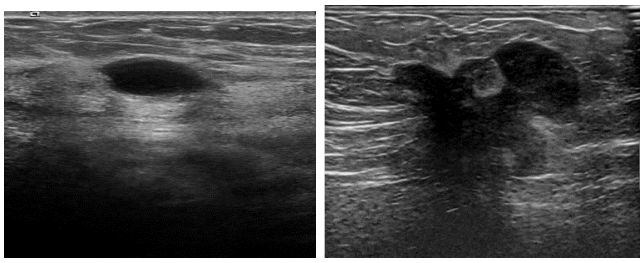


Fig. 1. begnin tumor malignant tumor

* PhD advisor: Dr.Domènec Savi Puig Valls, Dr. Mohamed AbdelNasser, and Dr. Miguel Angel

2 Problem Statement

Semantic image segmentation, which assigns per-pixel predictions of object categories for the given image, is a fundamental problem in computer vision. In the last years, many methods achieved an impressive result in the image segmentation problem, however the collection of labeled data for the task of semantic segmentation is expensive and time-consuming, as it requires dense pixel-level annotations. While deep CNNS based semantic segmentation approaches have achieved impressive results by using large amounts of labeled training data, their performance drops significantly as the amount of labeled data decreases [1].

Many deep learning architectures have been proposed to solve segmentation problem in the medical images like FCN, SegNet UNET, and GAN [2]. The UNET architecture [3] is the state-of-the-art in the medical image segmentation. UNET was the first architecture designed especially for medical image segmentation. It achieved a good result on the small dataset. UNET has encoder-decoder structure, which reduces the spatial dimension to extract features and then leverages up sampling to recover spatial extent. so it uses skip connections to preserve the spatial information, which is help in improving the segmentation task. The Encoder-Decoder architecture [3] is a neural network structure based on FCN improvements. The architecture is mainly composed of two parts, in which the encoder captures deep semantic information through several down-sampling processes; the decoder part gradually restores the space and detail information of the input image through several up-sampling operations. Recently, many deep learning based models have been proposed for breast tumor segmentation in BUS images, especially fully convolutional network (FCN) [4] and U-Net, have been successfully applied to this field and achieve outstanding performance for instance, Yap et al. [5] developed several FCN-based variants for the semantic segmentation of breast lesion in BUS images. With a dataset of 113 malignant and 356 benign BUS images, they achieved a dice score of 0.7626 on benign. Lesions whereas achieved 0.54 on malignant Lesions. Almajalid et al. [6] modified and improved U-Net for lesion segmentation based on the contrast enhanced and speckle-reduced BUS images. With a dataset of 221 BUS images, they achieved a dice score of 0.825.

However, breast tumor segmentation in BUS images segmentation remains an open problem due to the poor image quality and large variations in the sizes, shapes, and locations of breast lesion. In our research we used different semantic segmentations architectures to segment breast tumors in image. we compared the performance of different loss functions with different semantic segmentation models.

3 Methodology

In our research, we developed breast tumor segmentation models based on deep learning CNN models namely UNET and RESUNET with different loss functions. Specifically, we used various loss functions in order to check which one is more suitable to our data set and more effective, we chose to use them with UNET model, the data set used in this research is provided by UDIAT Diagnostic Centre of Sabadell, Spain. the size of the data set were 163 images with its grounds truth images. We divided the dataset as 113 images as train data and 50 images as test data, for training, we have used batch size of 100 and Adam optimizer with learning rate 0.0001 and with 20 epochs, and standard data-augmentation techniques (rotation range, width shift range, height shift range, shear range, zoom range, and horizontal flip) are applied, after we performed the data augmentation on the training dataset, the training dataset is increased to 2260 images. We have performed experiments using different loss functions, the loss functions can be defined as follow:

1. **Cross entropy** measure of the difference between two probability distributions for a given random variable or set of events

$$\mathrm{LBCE}\left(y, y^{\wedge}\right) = -\left(y \log\left({}^{\wedge}y\right) + (1-y)\log\left(1-y^{\wedge}\right)\right) \tag{1}$$

2. **Dice Coefficient loss** measure of overlap between the predicted sample and targeted sample, it's used for the binary data

$$\mathrm{Dice} \ = 1 - 2|A \cap B|/|A| + |B| \tag{2}$$

3. **Focal Loss** It is an improved version of Cross-Entropy Loss (CE) It down-weights the contribution of easy examples and enables the model to focus more on learning Hard examples. It works well for highly imbalanced class Scenario. So an extra parameter added (1- pt) to the cross-entropy loss, with a tunable focusing parameter 0. So focal loss can be defined as

$$\mathrm{FL(pt)} = -\alpha \mathrm{t}(1-\mathrm{pt})\gamma \log(\mathrm{pt}) \tag{3}$$

4. **Tversky loss** It's a generalization of Dice's coefficient. It adds a weight to FP (false positives) and FN (false negatives)

5. **Boundary with dice** Boundary loss, which takes the form of a distance metric on the space of contours, not regions. This can mitigate the difficulties of highly unbalanced problems because it uses integrals over the interface between regions instead of unbalanced integrals over the regions

4 Preliminary results and future work

We evaluated the segmentation performance of the proposed experiments are conducted on the UNET and ResUNET, and UNET outperformed the ResUNET since it achieved 0.823, whereas the ResUNET achieved 0.767

Fig. 2 shows the comparison results of the two approaches

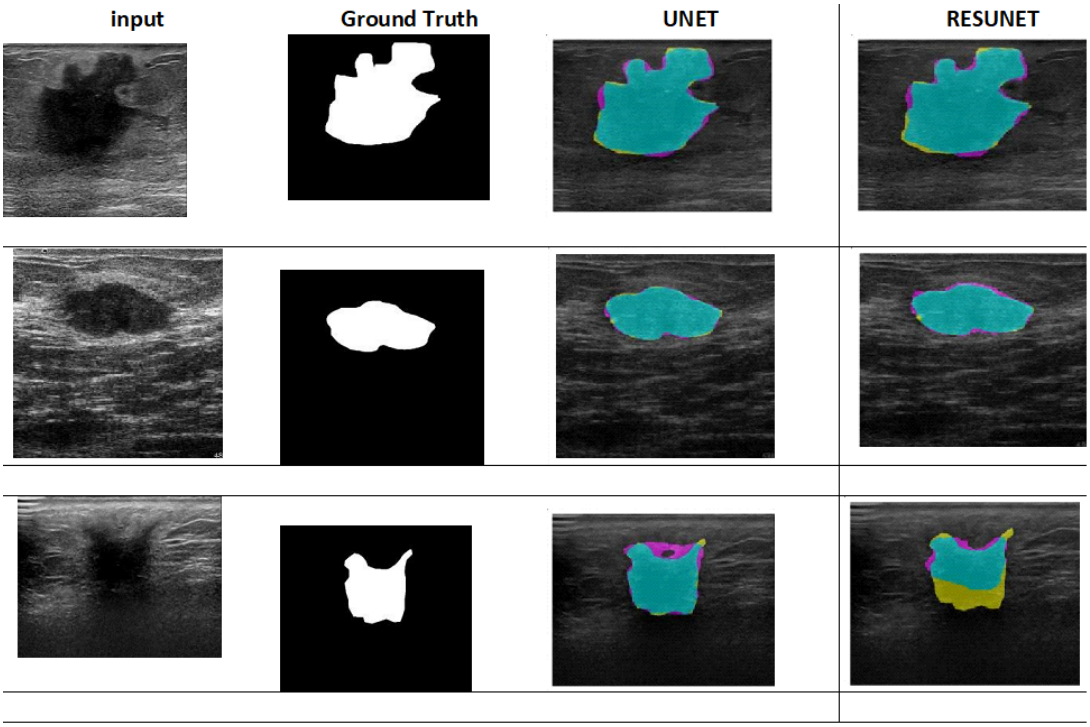


Fig. 2. Example of the predicted segment of the breast tumor for both UNET and ResUNET models Note: Cyan (TP) Red (FP) Yellow(FN) Background (TN)

We decided to evaluate the loss functions with UNET model since its outperformed the ResUNET model, the Table 1 shows the performance of the model with each loss function, the tversky loss function with tuned hyperparameter is outperformed the remaining mentioned loss functions

Methods	Accuracy	Dice	IOU(jaccard)	Sensitivity	Specificity
Cross entropy	0.980	0.823	0.721	0.663	0.997
DICE COEF	0.9976	0.849	0.738	0.806	0.997
Tversky (alpha=0.3,beta=0.7,smooth=1e-6)	0.9967	0.861	0.755	0.859	0.9972
Focal Loss alpha =0.25,gamma =2)	0.9970	0.826	0.707	0.797	0.9962
BCE+dice (took means of them)	0.9968	0.818	0.696	0.735	0.9982
boundary with dice	0.9956	0.805	0.674	0.725	0.9979

Table 1. Evaluation metrics of the UNET with various loss functions

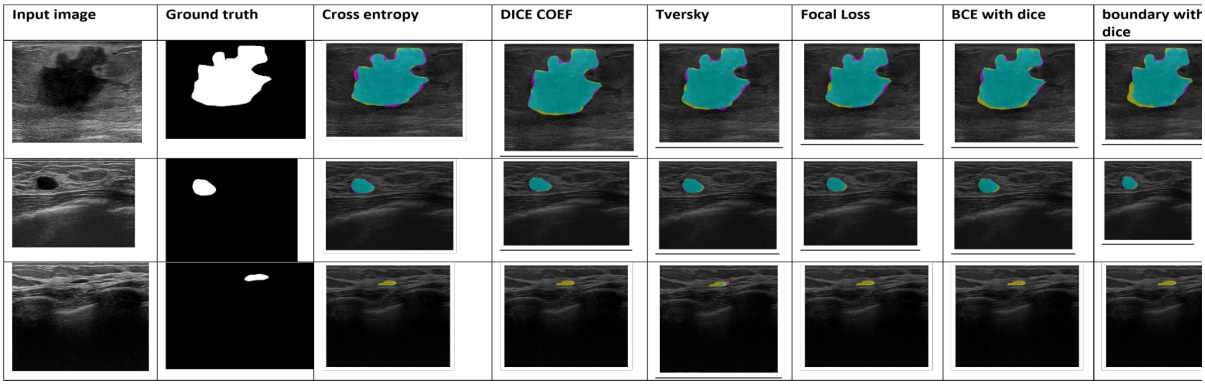


Fig. 3. Example of the predicted segment of the breast tumor of UNET model with various loss functions

In the future work we will perform FCDensenet model on our dataset, also we will evaluate FCDensenet model with Dual Attention, Dilated convolution and Multiscale contextual information.

References

[1] Zhao, Xiangyun et al. "Contrastive Learning for Label Efficient Semantic Segmentation." Proceedings of the IEEE/CVF International Conference on Computer Vision.

[2] Fu, Yabo, et al. "A review of deep learning based methods for medical image multi-organ segmentation." Physica Medica 85 (2021): 107-122.

[3] Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer assisted intervention. Springer, 2015, pp. 234–241.

[4] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence , vol. 39, no. 4, pp. 640–651, 2017.

[5] M. H. Yap et al., "Breast ultrasound lesions recognition: end-to-end deep learning approaches," Journal of Medical Imaging, vol. 6, no. 1, p. 011007, 2019

[6] R. Almajalid, J. Shan, Y. Du, and M. Zhang, "Development of a deep-learning-based method for breast ultrasound image segmentation," in International Conference on Machine Learning and Applications(ICMLA), 2018, pp. 1103–1108.

Optimizing the first convolutional layer

João Paulo Schwarz Schüler *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
joaopaulo.schwarz@estudiants.urv.cat

1 Introduction

In 1989, LeCun et al. [7] devised the first Convolutional Neural Network (CNN), which mimicked the organization of neural cells in the visual cortex as convolutional filters. This new type of neural network was able to recognize 10 digits in hand-written text very accurately. The majority of existing CNN models deal with the basic Red-Green-Blue (RGB) color values from input pixels. Despite this is the obvious choice taking into account that digital images are usually encoded with RGB, it's curious that very few researchers have attempted to train their networks on images encoded with other color spaces such as Hue-Saturation-Lightness (HSL) or CIE-LAB, the definition of which are vastly known and long-standing in the fields of color perception [2] and colorimetry [6]. The rationale behind trying other color spaces than RGB is based on evidences that the human color vision transforms the initial neural signals from cones and rods into an opponent color model [5], where several layers of neurons convert the Short, Medium and Large wavelength neural signals, loosely related to blue, green and red hues, into other neural signals. In regards to the human color perception [2], these opponent signals are further processed and converted into perceptual color components, named as Hue, Saturation and Lightness. There are several computational models that convert RGB into HSL-related components, for example, Smith's HSI [3] and Yagi's HSV [4].

2 Materials and methods

In order to check our hypothesis, we will perform image classification experiments on the CIFAR-10 dataset [1], which consists of 60k 32x32 RGB labelled images, belonging to 10 different classes: airplane, automobile, bird, cat, etc. These images are taken from natural and uncontrolled lightning environment,

* PhD advisors: Domènec Puig, Santiago Romaní and Mohamed Abdel-Nasser

contain only one prominent instance of the object to which the class refers, and the object may be partially occluded or seen from an unusual viewpoint. We aim to explore a simple CNN able to obtain a reasonable test accuracy (above 80%) in the CIFAR-10 image classification task. We compare its behavior (accuracy variation, patterns in first layer filter, etc.) from the basic RGB to the CIE-LAB encodings. These experiments were made with a Free Pascal based neural network API [8].

3 Experiments

As a baseline, we defined a single-branch CNN architecture small enough to classify CIFAR-10 dataset with at least 80% test accuracy. This single-branch architecture is shown in figure 1.

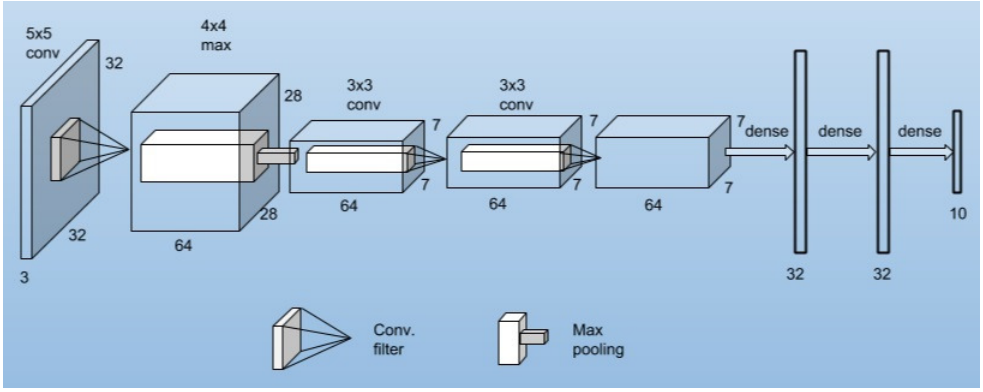


Fig. 1. Graphical representation of the single-branch baseline CNN architecture.

One of the purposes of our research is to create an architecture that takes advantage of separated chromatic and achromatic channels, which are readily available in color spaces such as CIE-LAB or HSV, as explained in the introductory section. To this aim, we propose to create two separate paths for the first convolutional layer, each one dedicated to each type of pixel information (achromatic/chromatic), in order to specialize the first layer filters of the CNN to the mentioned aspects of the scene (light variations, object boundaries). We hypothesize that this specialization may lead to a better object identification, as a consequence of a more object-related representation of the image content. Figure 2 shows the proposed two-branch architecture, where the top branch processes the single achromatic channel while the bottom branch processes the two chromatic channels. For example, we can convert RGB into CIE-LAB color encoding, hence the L channel is fed into the top branch, while the AB channels are fed into the bottom branch.

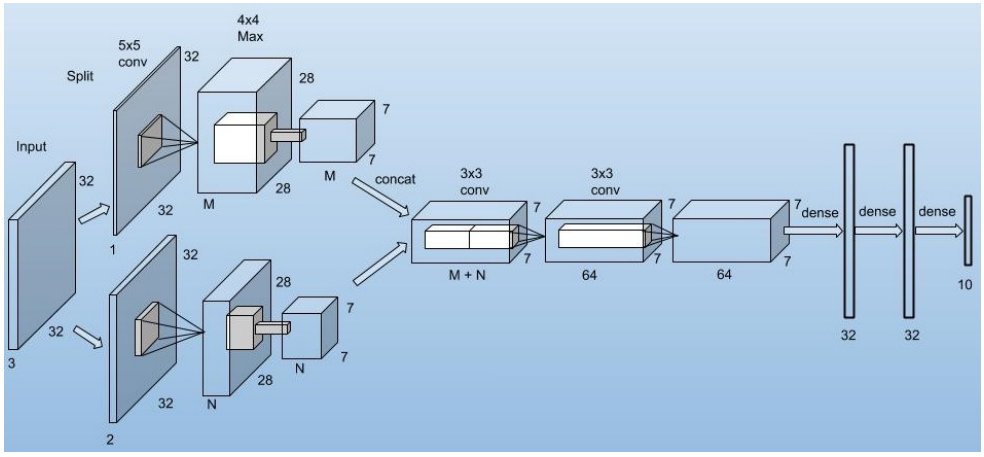


Fig. 2. Graphical representation of the single-branch baseline CNN architecture.

4 Results

As shown in the table 1, our baseline RGB model obtained 84.4% accuracy with 15.5 million floating point operations on the forward pass while our two-paths model obtained 84.7% accuracy with 11.7 million flops meaning a reduction about 29% in the required forward pass computation. The figure 3 shows the L (achromatic path) and the AB (chromatic path) learned patterns.

model	color space	accuracy	million flops
baseline	RGB	84.4%	16.5
two-paths	LAB	84.7%	11.7

Table 1. RGB baseline and LAB two-paths results.

5 Conclusions

By splitting LAB filter values into two branches, one for L and another for AB, we can force a CNN to find prototypical sets of achromatic/chromatic filters allowing the CNN to achieve similar accuracy while decreasing the required computation. In essence, we have devised a modification of the first layer of a CNN into two branches, which optimizes the number of weights when dealing with a color encoding that separates achromatic from chromatic channels, such as LAB, HSL, etc. Although the proposed architecture does not increase the validation accuracy significantly, it points out that uncorrelating the input features eases the learning task of any CNN. As a future work in this line, we aim to find out other "correlations" in mid-level or high-level layers, hence we may be able to specialize the network neurons to different types of information.

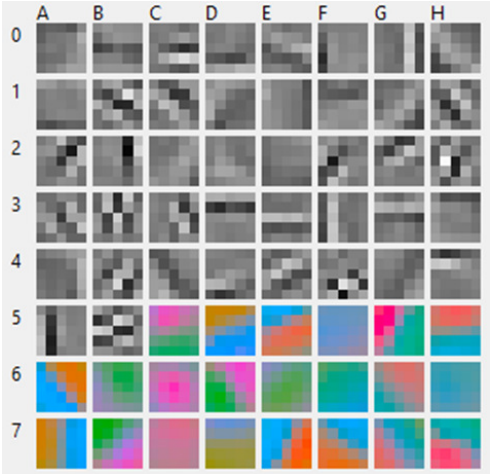


Fig. 3. Learned patterns in the L and the AB paths.

References

[1] A. Krizhevsky, G. Hinton. Learning Multiple Layers of Features from Tiny Images. 2009. doi:10.1.1.222.9220.

[2] A.R. Robertson. Color Perception. *Phys. Today.*, 45 (1992) 24–29. doi:10.1063/1.881324.

[3] A.R. Smith. Color Gamut Transform Pairs. *SIGGRAPH Comput. Graph.*, 12 (1978) 12–19. doi:10.1145/965139.807361.

[4] D. Yagi, K. Abe, H. Nakatani Segmentation of Color Aerial Photographs Using HSV Color Models. *Proc. IAPR Work. Mach. Vis. Appl. MVA*, December December 7-9, 1992, Tokyo, Japan, 1992: pp. 367–370. http://b2.cvl.iis.u-tokyo.ac.jp/mva/proceedings/CommemorativeDVD/1992/papers/1992367.pdf.

[5] E. Hering Outlines of a Theory of the Light Sense. *Harvard University Press*, 1920.

[6] G. Wyszecki, W. S. Stiles. Color Science: Concepts and Methods, Quantitative Data and Formulae, 2nd Edition *Color Sci. Concepts Methods, Quant. Data Formulae, 2nd Ed.*, Pp. 968. ISBN 0-471-39918-3. Wiley-VCH , July 2000. (2000). 2007.

[7] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel Backpropagation Applied to Handwritten Zip Code Recognition *Neural Comput.*, 1 (1989) 541–551.

[8] Schuler, J. P. S. CAI NEURAL API https://doi.org/10.5281/zenodo.5810077 https://github.com/joaopauloschuler/neural-api

Efficient Data Augmentation Techniques for Lesion Detection in Breast Tomosynthesis Images Using Deep Learning Models

Loay Hassan *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili Tarragona, Spain
loay.abdelrahimosmanhassan@urv.cat

1 Introduction

Breast cancer is one of the most common malignancies in women worldwide and a leading cause of death [1]. On the other hand, early diagnosis has been repeatedly shown to reduce overall disease burden and mortality and help to get successful treatment. The classical imaging diagnostic tools used for breast cancer screening are mammography (X-ray images of the breast) and breast ultrasonography (BUS). However, the 2-D imaging modality of these images causes the presence of high breast density (dense fibro glandular tissue in the breast) which limits the sensitivity and specificity of breast lesion detection [2].

Digital breast tomosynthesis (DBT) which is a new 3-D breast cancer screening technique, has the ability to address the limitation of tissue overlapping and superimposition in mammography [3] by providing superior tissue visualization which yields enhanced breast lesion detection rate. However, as the number of slices to evaluate grows, physicians' oversight of findings increases which creates clinical workflow challenges since it is necessarily that radiologists are required to examine a greater number of slices per breast volume. As a result, computer-aided detection (CAD) is regarded as the ideal solution for clinical DBT and plays a greater clinical role in improving work performance than traditional digital mammography. Furthermore, Due to the higher mass margin visibility in DBT images, it is also probably that CAD will perform better than with mammographic images [4].

Despite the fact that many automated lesion detection approaches for accurately detecting breast cancer in mammographic images have been proposed in the literature, alongside the lack of enough annotated DBT images which held back the number presented detection methods for DBT, breast cancer detection in mammographic and DBT images is still a challenging task. In this

* PhD advisor: Mohamed Abdelnasser and Domenec Puig

Work, we present an automated deep learning-based breast lesion detection method for DBT images based on investigating the impact of two data augmentation techniques called channel-replication and chancel-concatenation in improving the breast lesion detection results of robust object detection models like YOLO [5] and Faster R-CNN [6].

2 Deep Learning Based Breast Cancer Detection System

Deep learning is a part of machine learning that has revolutionized the area of computer vision and has been employed in various of medical detection application including breast cancer detection. The key elements of our proposed method are data augmentation, deep learning based detector, and non-maximum suppression (NMS).

2.1 Data augmentation

Data augmentation techniques are used in deep learning by implementing different image manipulation algorithms to increase the number of training images. In this work, We analyze the effect of two different data augmentation techniques.

- **Channel-replication.** In this practice, the N training images are increased by $6N$ through flipping all images in the training set horizontally, then gamma correction is applied for each image I_s (original and flipped image) following the Equation 1 to adjust the overall brightness of an image to generate I_γ. In addition,I_{clahe} images is generated by applying the contrast limited adaptive histogram equalization (CLAHE) to enhance the image Local Contrast [7]. To calculate the clip limit for the CLAHE algorithm, we follow Equation 2.

$$I_\gamma = 255 \times (\frac{I_s}{255})^\gamma \tag{1}$$

  Where I_γ is the output image for gamma correction and γ is the gamma correction factor.

$$cliplimit = \frac{W \times H}{L} \left(1 + \frac{\alpha}{100} \left(S_{\max} - 1\right)\right) \tag{2}$$

  Where $W \times H$ is the number of pixels in each histogram calculated region, L is the number of gray-scales, α is a clip factor, and S_{max} is the maximum allowable slope.
- **Channel-concatenation.** In this practice, unlike the traditional augmentation techniques, the number of data is not increased. But, a new 3-channel training images (I) has been produced by concatenating the original image with two post-processed images as shown in Equation 3,

following the idea in [8]. The two filtered images (I_γ with $\gamma = 0.5$ and I_{clahe} with $\alpha = 1$) is concatenated with the original gray-scale image I_g.

$$I = Concat(I_g, I_\gamma, I_{clahe}) \tag{3}$$

Here, I,I_g,I_γ, and I_{clahe} is output image, image after gamma correction and image after CLAHE equalization, respectively.

2.2 Deep learning based competent detection models

We used two widely known and efficient deep learning-based object detectors: YOLO [9] and Faster-RCNN [6], to develop the individual deep learning-based detection models.

In this work, we employed YOLO Version 5, which is now the most advanced object detection algorithm of the YOLO family available. It is a novel approach that detects objects in real-time with great accuracy. It uses a single neural network based on convolutional neural network (CNN) to process the entire image then separates it into parts and predicts bounding boxes and probabilities for each component. YOLOv5 is available in four models, namely (YOLO-Small (S), YOLO-Medium (M), YOLO-Large (L), and YOLO-XLarge (XL)).

In addition, faster R-CNN detector is employed for further attestation of the proposed approach. Faster R-CNN the most widely used state of the art version of the R-CNN family. It comprises four major parts: 1) a feature extractor stage–usually using a CNN, 2) a region proposal (RPN) algorithm which utilise a CNN network instead of using a selective search algorithm to predict bounding boxes of possible objects in the image with a confidence score that yield accelerating training time and improving feature representation, 3) a classification layer to predict which class this object belongs to, and 4) a regression layer to make the coordinates of the object bounding box more precise.

2.3 Implementation

Firstly, the DBT images dataset was divided patient-wise into training and testing sets. During the training phase, we use the data augmentation techniques described in section 2.1 to generate two training sets (training set by channel-replication augmentation and training set by channel-concatenation augmentation). Then, we train each of the detectors mentioned in section 2.2 individually for each of these training sets.

Second, the trained models are used to predict bounding boxes for each DBT image in the test set during the testing phase. A single bounding boxes list contains all predicted bounding boxes from a single DBT image is passed to the (NMS) algorithm, which selects the best bounding box from a set of overlapping or duplicated boxes.

2.4 Experimental results

Table 1 presents a quantitive comparison between the four YOLOv5 models (YOLO-S, YOLO-M, YOLO-L and YOLO-XL) and the faster R-CNN model trained on the both training dataset produced by channel-replication augmentation and channel-concatenation augmentation in terms of true positive rate (TPR), F1-score, and mean average precision–mAP (IoU threshold = 5).

Table 1: The performance of the deep learning detection methods [10]

Augmentation	**Channel-replication**					**Channel-concatenation**				
Models	**YOLOv5**				**Faster R-CNN**	**YOLOv5**				**Faster R-CNN**
	S	**M**	**L**	**XL**		**S**	**M**	**L**	**XL**	
TPR	38.8	31.8	24.2	22.7	50	47	39.4	39.4	47	56.1
F1-Score	48.5	45.7	36.1	35.7	54.1	52.5	51.7	56.6	51.4	57.4
mAP [iou = 0.5]	31.8	34.1	26.2	26.7	45.1	48.7	38.9	40.4	41.8	46.8

As one can see, YOLO-S achieved the best lesion detection results when compared to the other YOLO models for channel-replication. However, faster R-CNN could be more suitable for DBT images as it has more promising breast lesion detection results that surpassed all YOLO models on all measures. It is notable that training the deep learning detectors based on channel-concatenation yields noticeable improvements on all metrics [10], where the performance of the YOLO-S model increased by 17 points in terms of mAP. Besides, the TPR and F1-score of the faster RCNN were also advanced 6% and 3.3%, respectively. On the basis of the above analysis, we can conclude that channel-concatenation data augmentation technique can significantly improve the breast lesion detection results for deep learning-based breast lesion detectors like YOLO models and faster R-CNN.

3 Conclusions

In this work, we present the strength of two data augmentation strategies (channel-replicate and channel-concatenation) while building state of the art breast lesion detection models based on deep learning for digital breast tomosynthesis.

The study demonstrate that applying the channel-concatenation data augmentation strategy helps improve the detection accuracy of all deep learning models. With a publicly available digital breast tomosynthesis dataset. The future work will be focused on the development of a lesion detection approach based on the combination of robust deep learning-based detectors.

Acknowledgement. The Spanish Government partly supported this research through Project PID2019-105789RB-I00.

References

[1] H.D. Cheng, Juan Shan, Wen Ju, Yanhui Guo, and Ling Zhang. Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern Recognition*, 43(1):299–317, 2010.

[2] Emine Devolli-Disha, Suzana Manxhuka-Kërliu, Halit Ymeri, and Arben Kutllovci. Comparative accuracy of mammography and ultrasound in women with breast symptoms according to age and breast density. *Bosnian journal of basic medical sciences*, 9(2):131–136, May 2009.

[3] Mark A. Helvie. Digital mammography imaging: breast tomosynthesis and advanced applications. *Radiologic clinics of North America*, 48(5):917–929, Sep 2010.

[4] Ming Fan, Yuanzhe Li, Shuo Zheng, Weijun Peng, Wei Tang, and Lihua Li. Computer-aided detection of mass in digital breast tomosynthesis using a faster region-based convolutional neural network. *Methods*, 166:103–111, 2019. Deep Learning in Bioinformatics.

[5] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, 2016.

[6] 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.

[7] Kitti Koonsanit, Saowapak Thongvigitmanee, Napapong Pongnapang, and Pairash Thajchayapong. Image enhancement on digital x-ray images using n-clahe. In *2017 10th Biomedical Engineering International Conference (BMEiCON)*, pages 1–4, 2017.

[8] Moi Hoon Yap, Manu Goyal, Fatima Osman, Robert Martí, Erika Denton, Arne Juette, and Reyer Zwiggelaar. Breast ultrasound region of interest detection and lesion localisation. *Artificial Intelligence in Medicine*, 107:101880, 2020.

[9] Glenn Jocher. Yolov5 Pytorch Implementation. `https://github.com/ultralytics/yolov5`. Accessed: 2021-05-29.

[10] Loay Hassan, Mohamed Abedl-Nasser, Adel Saleh, and Domenec Puig. Lesion detection in breast tomosynthesis using efficient deep learning and data augmentation techniques. In *Frontiers in Artificial Intelligence and Applications*. IOS Press, October 2021.

Road Damage Detection Using Yolov5

Ammar Mohammed Okran ⋆

Department of Computer Engineering and Mathematics, URV, Tarragona, Spain
ammar.okran@urv.cat

1 Introduction

Road damage detection is one of the most important issues related to safety, which is directly related to human life and vehicles. Most of the basic infrastructure of most countries dates back to previous decades. For example, a country like Japan, during the boom of economic growth in the late 20^{th} century, extensively built roads, bridges, etc[1]. That is, the infrastructure age is now more than 50 years old, and needs to be inspected and proper maintenance conducted.

The process of road inspection and maintenance is time-consuming and costly, since this infrastructure extends for thousands of kilometers, and to detect damaged parts by traditional methods, requires advanced survey equipment, huge financial resources, and experts. For this reason, most municipalities neglect the detection procedures[2]. The problem of aging infrastructure is prevalent in other countries such as the United States of America[3], and it is considered a vexing problem for municipalities. However, the need for efficient and advanced ways to maintain infrastructure has become an urgent necessity.

Recently, several methods and studies have been conducted to address this problem including methods of using laser technology or image processing, in addition, using neural networks and machine learning techniques. In 2018, the Road Damage Dataset 2018[4] was published and a challenge was held in Seattle, USA, based on this dataset. A total of 59 teams from 14 countries participated in this competition, all the top results use an ensemble that applies multiple NNs.

Our work is aimed to detect and classify the road damages in order to facilitate decision conducting for road managers to do a proper maintenance according to the damage type. To do this, Yolov5 is used in our experiments due to its robustness and promising results as well as the Road Damage Dataset 2018.

⋆ PhD advisors: Domènec Puig, Mohamed Abdelnasser and Hatem Abdellatif.

2 Methodology

2.1 Dataset

The dataset consists of 9053 labeled road damage images with the resolution of 600X600 pixels, which are taken from different cities in Japan (Adachi, Chiba, Ichihara, Muroran, Nagakute, Numazu, and Sumida). In this dataset we have 8 classes of road damages separated as follows: D00, D01, D10, D20, D40, D43, and D44. These classes are defined in the table1. In these 9,053 images there are 15,435 instances are distributed to all 8 classes as shown in the figure1

Damage type			Detail	Class name
Crack	Linear crack	Longitudinal	Wheel mark part	D00
			Construction joint part	D01
		Lateral	Equal interval	D10
			Construction joint part	D11
	Alligator crack		Partial pavement, overall pavement	D20
Other corruption			Rutting, bump, pothole, separation	D40
			Crosswalk blur	D43
			White line blur	D44

Table 1: *Source*: Road Maintenance and Repair Guidebook 2013 (JRA, 2013) in Japan. *Note*: In reality, rutting, bumps, potholes, and separations are different types of road damage, but it is difficult to distinguish these four types using images. Therefore, they were classified as one class, namely, D40.

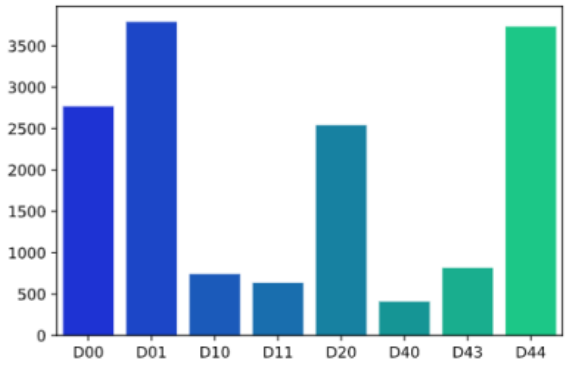


Fig. 1: The instances of each class in the Road Damage Dataset 2018.

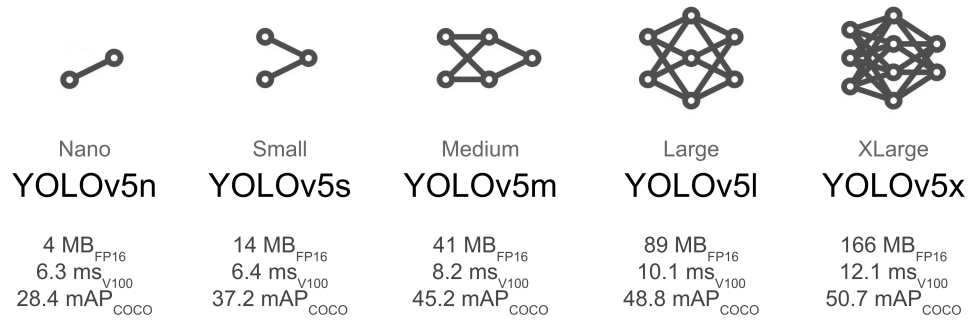


Fig. 2: YOLOv5 Models Comparison.

2.2 Train models using Yolov5

Yolov5 has multiple varieties of pretrained models as shown in the figure2 trained on the COCO dataset. In our experiments, we started the training process with YOLOv5x pre-trained model as initial weights with different image sizes (448 and 608) and with YOLOv5x6 pre-trained model as initial weights with image sizes (448 and 576), all with the default hyperparameters.

3 Results

Figure3 shows the predictions results compared to the ground truth, the best model model we got achieved an F1-score of 0.631, where this result without applying any improvements such as test-time augmentation or model ensembling. Applying the test time augmentation and model ensembling led to improving the predictions almost in all metrics, table2 shows the results in details.

Yv5x_448	Yv5x_608	Yv5x6_576	TTA	Precision	Recall	mAP@.5	mAP@.5:.95	F1-score
X				0.644	0.617	0.64	0.364	0.63
X			X	0.634	0.614	0.633	0.361	0.623
	X			0.629	0.633	0.625	0.359	0.631
	X		X	**0.695**	0.608	0.647	0.374	**0.648**
		X		0.617	0.644	0.642	0.364	0.63
		X	X	0.613	0.657	0.65	0.37	0.634
X	X		X	0.631	0.649	0.658	0.378	0.639
X		X	X	0.59	**0.675**	0.664	0.376	0.629
	X	X	X	0.648	0.641	0.662	0.378	0.644
X	X	X	X	0.629	0.657	**0.666**	**0.381**	0.642

Table 2: The results for all trained models

(a) Labels of a batch of 16 images

(b) Predictions of a batch of 16 images

Fig. 3: Comparison between the Groundtruth and prediction

4 Conclusions

The results obtained using Yolov5 show that our approach was able to achieve results close to the state_of_the_art, where we can get a *mAP@.5* up to **0.666**, *mAP@.5:.95* up to **0.381** and *F1* score up to **0.648**.

Acknowledgement. This work has been funded by the Govt. of Spain under contract TIN2016-77836-C2-1-R and the Govt. of Catalonia as Consolidated Research Group 2017-SGR-688.

References

[1] Ministry of Land, Infrastructure, Transport and Tourism, White Paper on Present state and future of social capital aging. (2016) *Infrastructure maintenance information, (in Japanese).*, 2016.

[2] Tomiyama, K., Kawamura, A., Fujita, S. Ishida, T. (2013), An effective surface inspection method of urban roads according to the pavement management situation of local governments, *Journal of Japan Society of Civil Engineers, Ser. F3 (Civil Engineering Informatics),*69(2), I-54–I-62.

[3] AASHTO (2008), Bridging the Gap–Restoring and Rebuilding the Nation's Bridges, *American Association of State Highway and Transportation Officials, Washington DC.*.

[4] Maeda, Hiroya and Sekimoto, Yoshihide and Seto, Toshikazu and Kashiyama, Takehiro and Omata, Hiroshi, Road damage detection and classification using deep neural networks with smartphone images. *Computer-Aided Civil and Infrastructure Engineering*, 2018.

This book contains the proceedings of the 7th Doctoral Workshop in Computer Science and Mathematics - DCSM 2022. It was celebrated in Universitat Rovira i Virgili (URV), Campus Sescelades, Tarragona, on March 31, 2022. The aim of this workshop is to promote the dissemination of ideas, methods, and results developed by the students of the PhD program in Computer Science and Mathematics from URV.

Departament d'Enginyeria [DEIM] Informàtica i Matemàtiques
Universitat Rovira i Virgili
Escola Tècnica Superior d'Enginyeria
Universitat Rovira i Virgili