An enhanced UNet variant for Effective Lung Cancer Detection

Lerina Aversano
Dept. of Engineering
University of Sannio
Benevento, Italy
aversano@unisannio.it

Mario Luca Bernardi
Dept. of Engineering
University of Sannio
Benevento, Italy
bernardi@unisannio.it

Marta Cimitile
Dept. of Law and Economics
UnitelmaSapienza University
Rome, Italy
marta.cimitile@unitelmasapienza.it

Martina Iammarino
Dept. of Engineering
University of Sannio
Benevento, Italy
iammarino@unisannio.it

Chiara Verdone
Dept. of Engineering
University of Sannio
Benevento, Italy
chiverdone@unisannio.it

2022 International Joint Conference on Neural Networks (IJCNN) | 978-1-7281-8671-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/IJCNN55064.2022.9892757

***Abstract*—Lung cancer is one of the diseases with the highest mortality rate and early detection is key to making the disease as treatable as possible. The most common and useful method for recognizing pulmonary nodules is computed tomography, which allows them to be located and monitored. The disadvantage of this technique is that the scans have to be interpreted by doctors, who could make mistakes. The use of pulmonary CAD is therefore becoming increasingly widespread, a system capable of automatically analyzing CT images and providing information on possible suspicious regions found in the images. These systems, by offering radiologists a list of already marked regions of interest to view with particular attention, increase the efficiency of detection of small nodules and reduce reporting times by physicians. This study aims to accurately detect the location of pulmonary nodules through a Deep Learning approach with the use of computed tomography scans. In particular, it proposes the use of a new variant of the UNet architecture, called GUNet3++, which has been compared with the other types of this network. To validate the approach, the public LIDC-IDRI dataset was used, which collects pulmonary CT images of about a thousand patients with different types of cancer. The results obtained are very promising, showing a performance improvement compared to other UNet networks.**

***Index Terms*—Deep Learning, U-Net, CT scan Images, Lung Cancer**

I. Introduction

Lung cancer is a very serious disease that affects a vital organ for the human being, and unfortunately, it represents the first cause of death from cancer in industrialized countries [18]. Its incidence is constantly and continuously increasing [19], therefore early diagnosis can make a difference by allowing prompt treatment. In this regard, it is of fundamental importance to subject high-risk individuals to regular screenings [8], [20], thanks to which it is possible to detect cancer at a very early stage. There are several tests that can determine the health of the lungs, but the test considered most effective at an early stage is known as low-dose CT (computed tomography). The radiologist analyzes the images produced to look for abnormal areas in the lungs that could indicate cancer. A spot visible in the image is known as a "pulmonary nodule" and needs to be monitored. The problem is represented by the criticality of the interpretative analysis by the doctor who finds himself evaluating a high number of images with the risk of making mistakes and causing incorrect diagnoses. This explains the introduction of some computerized image analysis systems, called CAD (Computer-Aided Detection), which provide the localization and automatic identification of the pulmonary nodule. This method can significantly reduce errors, reducing time [12], and identifying half of the cases lost to humans [31]. In recent years, several CAD systems have been developed for the detection and classification of pulmonary nodules in CT [13], [30] imaging systems, some based on machine learning [32] and others on Deep Learning (DL) approaches [22]. Among the latter, which show better performance than traditional lung nodule detection systems, an example is the use of CT scan images for pulmonary nodule detection [1]. In particular, in [4] lobe-driven CT image clustering is used for classification.

Like the previous ones, this study also proposes a DL-based approach to automatically and accurately detect lung cancer by performing the semantic segmentation of the CT scans. More specifically, we use a generalized UNet-based architecture, called GUNet3++, that includes multiscale and dense skip connections to build a more accurate model of the detected lung nodules.

The document is structured as follows: in Section II

978-1-7281-8671-9/22/$31.00 ©2022 IEEE

the most relevant related works are reported, Section III concerns the pre-trained neural networks adopted, some fundamental concepts are in fact explained, and Section IV describes the approach used in detail. The description of the experiment is in Section V, and Section VI shows the results obtained. Finally, Section VII and VIII report threats to validity and conclusions, respectively.

II. Related Work

In recent years, Machine Learning and Deep Learning (DL) techniques have been increasingly adopted to deal with mining and classification problems in the medical field [2], [3], [10].

In their work [33], Zhang et al. present a very detailed systematic review of existing techniques that aim to diagnose nodules using CT images. This review shows that many approaches use 2D CNN, DBN, and MTANN [27], [28], although these are superficial and have daunting performance.

Starting from this assumption and given that the DL has a better classification capacity than the classical lung cancer detection techniques [9], the next step in the development of the DL techniques is the optimization of network performance [14], [25].

In the work [17], the authors combine DL techniques with additional information, such as clinical factors, in order to reduce possible classification errors; the ultimate goal is early diagnosis.

To achieve the same goal, Gao et al. [11] instead use the deep neural networks CNN, DNN, and SAE, on the LIDC-IDRI database [16], obtaining an accuracy equal to about 85%.

The study proposed in [35] aims at the classification of nodules from CT scans, but it uses the DeepLung approach and is based on two different phases. The first phase concerns the detection of candidate nodules based on a 3D Faster R-CNN and a second phase in which the deep characteristics are extracted and classification is carried out through a DPN network.

In the work [1], the authors aim to specialize classifiers on a specific lobe of the lung using an overall approach based on image grouping. Different pre-trained convolutional neural networks were used, the best classifier was selected for each. The resulting accuracy in the case of the classification of the single image is equal to 96%, while in the classification of the patient it is equal to 94%. In case a small number of images is available, it is possible to use approaches based on U-Net [23] which manage to obtain good performances. Therefore, there are several works based on the segmentation of images [6], [26]. Note that performance may vary based on the size of the receptive field in the convolution kernel. Unfortunately, the use of U-Net networks has limitations, the fixed receptive field of the kernel causes a poor diversity of characteristics, thus semantically different characteristic maps are obtained. The proposed approach aims to overrun these limitations by improving the semantic segmentation performance with a more effective feature extraction at both intra and inter-resolutions levels.

III. Background

This study aims to perform the semantic segmentation of CT scans to identify lung cancer lesions.

Semantic segmentation is the process of partitioning an image into meaningful regions where each pixel in a region is similar to the others in the same region in some properties or characteristics. The attention to the categorization of the pixels is very relevant as a support in the medical field as it represents a diagnostic tool, for example, a nodule is identified by pixels belonging to the same class [21].

The proposed approach is based on variants of the UNet network, called GUNET3++, therefore we briefly illustrate the main variants of the UNet network that led to the creation of the network used.

A. UNET

UNet is a fully convolutional [24] neural network, whose architecture has been developed to be able to work with fewer training images and produce more precise segmentation, even in the 3D imaging segmentation field [5].

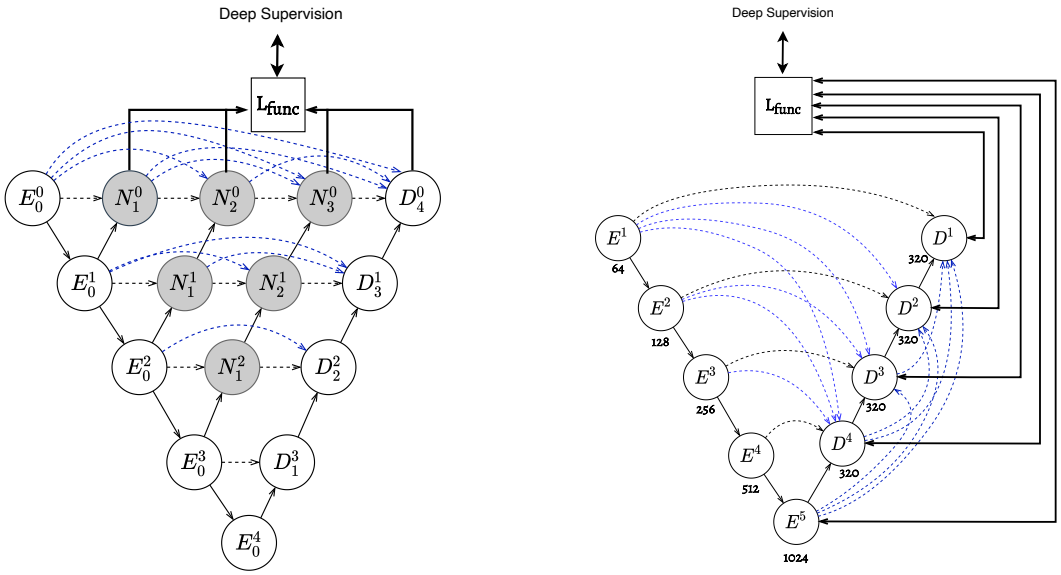
The idea behind UNet is to add successive layers to a standard contractual network where upsampling operations will take place, increasing the resolution of the output. This way a subsequent convolutional layer takes information on how to assemble the next output.

It has a U-shaped architecture and consists of a contracting path and an expansive path. The first is a typical convolutional network in which each convolution is followed by a rectified linear unit (ReLU) and a max-pooling operation. Therefore, the information on the characteristics is increased while the spatial information is reduced. The second path has the task of combining spatial information and characteristics through convolutions that use high-resolution characteristics deriving from the contraction path.

The network uses only the valid part of each convolution with no fully connected layer. Edge pixels are predicted based on the input image. This architecture is suitable for working with large images and allows you to improve image resolution by overcoming the limit of GPU memory.

B. UNET+ and UNET++ variants

Zhou et al. in [34] have introduced three different variants of the UNet network. The first is *UNet^e*, made up of an overall architecture made up of UNets of different depths that partly share the same codifiers while each has its decoder. This implies that the following networks do not supervise the decoders of the previous



(a) UNet++ variant

(b) UNet3+ variant

Fig. 1: Variants of UNet

ones, and the skip connections combine the maps of the characteristics of the decoder only in the same resolution scale resulting in too restrictive. The problem of skip connections has been solved in the second version, *UNet+*, where skip connections become direct between two adjacent nodes. This also solves the problem of no supervision signal between deep and shallow decoders. The presence of direct skip connections allows training the UNet + network both in a classic way and with in-depth supervision. Finally, *UNet++* is constituted starting from UNet+ connected to the decoders through densely skip connections, resulting in a more flexible fusion of functionality at the decoder nodes. Each node of the decoder performs two different operations: it horizontally combines the characteristics at the same resolution from all the previous nodes and vertically integrates the multi-scale functionalities on different resolutions of the previous nodes. In this way, maps of the aggregate characteristics are obtained that generate a more robust training process, less loss of semantic information, and greater accuracy. In Figure 1-(a), we report the architecture scheme in which the decoder nodes are indicated with D, the encoders with E, and those of the pyramidal block with N. The diagram shows how nodes of level (l) greater than 1 receive $l+1$ input. Of these inputs, l corresponds to the outputs of the previous levels, while $(l+1)^{th}$ corresponds to the skip connection. Also, in the figure, the skip connections are shown in blue.

C. UNET3+ variant

UNet3+ is another variant that derives from UNet, which has the particularity of exploiting the deep supervision extended on several scales for the training phase, and modifying the skip connections by combining the multiscale functionalities. It uses a different approach than previous versions as it uses information from all available scales. Each node of the decoder is capable of acquiring maps of the characteristics of the same scale, of a higher and lower scale. In this way, the deep and superficial semantic data are combined using a feature aggregation mechanism and it is possible to obtain coarse and fine-grained semantic information on all the resolution scales analyzed. The peculiarity of this architecture is shown in Figure 1-(b), where it is possible to notice that the network, on all available scales, learns both fine-grained and coarse-grained information. The decoder, at each level, acquires maps of the same scale from the encoder, those of the smallest characteristics and those of the largest characteristics of the decoder.

IV. The GUNet3++ approach

This work aims at the precise and automatic detection of lung nodule position using CT images. The approach is based on DL, in particular, the GUNet3++ network is used, for the CT-scan semantic segmentation.

The new variant of the UNet network proposed is a compromise between the UNet++ and UNet3+ networks and is called GUNet3++. As already mentioned in the previous section the UNet++ network adds dense skip connections on each scale, while UNet3+ performs an aggregation operation of multiscale functionality on each decoder. The GUNet3++ architecture was born as a combination of the two networks mentioned and includes the strengths of each. As can be seen from Figure 2,

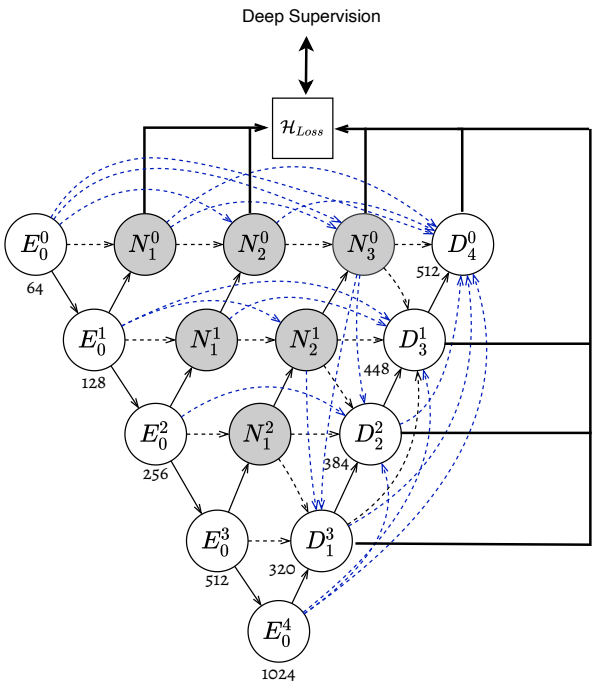


Fig. 2: The GUNet3++ architecture.

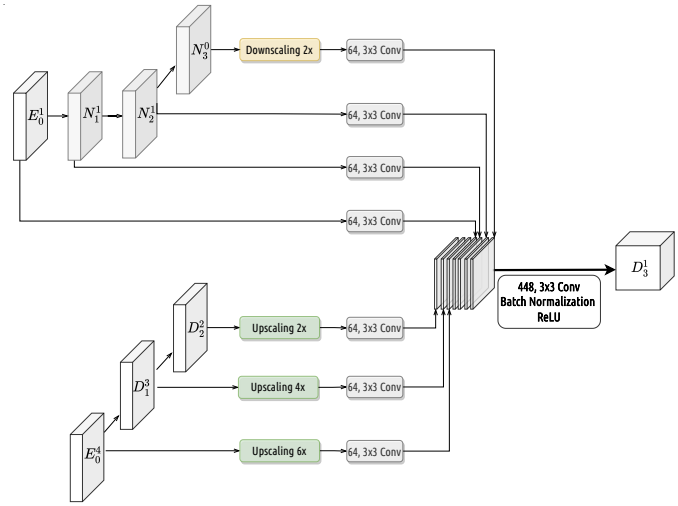


Fig. 3: Aggregated features map for decoder D_3^1

the architecture is able to propagate information at any scale from shallow to deeper nodes through the dense pyramidal block of transducers. In addition, the network is also able to learn from different regions thanks to the multiscale skip connections.

As an example of possible operation of the decoder, we show the Figure 3, where we want to generate the output signal to the decoder node D_3^1 through upscaling and downscaling. A set of six feature maps were used to achieve the output in this case. The resulting feature map is 448 filters obtained by convolution with 64 filters of size $3x3$. The figure shows that, with the exception of E_3^0, N_1^1, and N_2^1, all channels are upscaled or downscaled.

Defining i as the down-sampling layer on the encoders path, j as the convolution layer across the skip connection, and $N+1$ as the number of scales, the output of the generic node $Y^{i,j}$ is defined as:

$$y^{i,j} = \begin{cases} H_c\left(\left[[y^{i,k}]_{k=0}^{j-1}, U(y^{i+1,j-1})\right]\right) & \substack{i\in[0,N/2]\\ j\in[1,N-i-1]} \\ H_c\left(\left[[y^{i,k}]_{k=0}^{N-i-1}, C\big(D\big(y^{i-(k+1),j+k}\big)\big)_{k=0}^{N-1-j}, \right.\right. & \substack{i\in[0,N-1]\\ j=N-i} \\ \left.\left. \quad C\big(U\big(y^{i+(k+1),j-(k+1)}\big)\big)_{k=0}^{N-(i+1)}\right]\right) & \\ H_c(D\left(y^{i-1,j}\right)) & \substack{i>0\\ j=0} \end{cases}$$

where $C(\cdot)$ is the convolution operation, $H_c(\cdot)$ is the convolution followed by the activation operation, $D(\cdot)$ is the down-sampling operation, $U(\cdot)$ represents the up-sampling operation, and $[\,\cdot\,]$ is the concatenation operator.

The network was trained using a hybrid loss function that takes into account both the soft dice coefficient loss and the pixel-wise cross-entropy loss. Like the other variants, it is possible to carry out deep supervision on several scales in order to learn the hierarchical representations from the maps of the characteristics of all the resolutions. For this reason, the loss function $\mathcal{G}_{Loss}$ is defined as a weighted sum of an hybrid loss of each decoder:

$$\mathcal{G}_{Loss} = \sum_{k=1}^{N_D} \mathcal{D}_{Loss}(G, Y^k)$$

where k is an index across the N_D decoders, (G, Y^k) are the true labels and the labels evaluated by decoder k for every pixels in the batch, and $\mathcal{D}_{Loss}(G, Y^k)$, defined as the sum of pixel-wise cross-entropy and dice-coefficient contribution, can be fomulated as:

$$\mathcal{D}_{Loss}(G, Y^k) = -1/\mathcal{P}_S \sum_{c=1}^{\mathcal{N}_C} \sum_{n=1}^{\mathcal{P}_S} \left(2 * \frac{\mathcal{L}(p_n, c) * y(p_n, c)}{\mathcal{L}(p_n, c)^2 + y(p_n, c)^2} + \mathcal{L}(p_n, c) * \log(y(p_n, c))\right)$$

where, c is one among the $\mathcal{N}_C$ classes, p_n are the $\mathcal{P}_S$ pixels belonging to the current batch, $\mathcal{L}(p_n, c) \in G$ are the target labels, $y(p_n, c) \in Y^k$ are the predicted labels by the decoder k.

V. Experiment Description

In this section, we describe the adopted CT scans dataset and the details on the implementation and validation of the proposed model.

A. Dataset

The dataset adopted is the Lung Image Database Consortium image collection (LIDC-IDRI) [16], an international resource accessible from the web. It contains diagnostic and screening scans of lung cancer with chest CT from 1010 patients. For each image, a double-check was carried out. In the first phase, each doctor independently examined and marked the lesions, in the second phase the doctors compared their evaluations and expressed

a final opinion. More specifically, a nodule is indicated when everyone agrees on the existence of a lesion whose diameter exceeds 3 mm. In addition, for each patient image, there is also a metadata file that records the results of an image annotation process performed by the four expert radiologists. The annotation is aimed to identify the nodule position in the image. It is described using the metric Nodule Contour ROI, giving a complete three-dimensional contour of the nodule [16], thanks to which it is possible to obtain a mask for each representative image of the nodules present, and an empty mask for images where it is not indicated the presence of nodes. Scans that had low quality and all images belonging to patients who did not have enough scans were removed from the original dataset.

Therefore, three datasets were built, the Oracle Nodules containing the 32,606 images of the LIDC-IDRI database of 500 patients with related metadata, with BMP format and resolution of 512×512 pixel (i); the CT scan training data set containing the same images as the previous one, of which 28,227 do not contain nodules (ii); and the CT scan test dataset containing images of the remaining 500 patients (iii). Notice that the masks obtained by the ROI metrics are necessary for the training of the UNET network. However, the CT scans and the corresponding masks are used to train the network to perform the semantic segmentation of the nodules.

B. Experimental setting

The experiment conducted had a dual purpose, to detect the region of interest of the lung lesion in CT scans with the highest possible accuracy rate, through an approach based on the use of a GuNet3++ network, and to use window correction to evaluate the performance in identifying nodules in the different images.

The experimentation was conducted thanks to the use of an Intel Core i9 7920X with 18 cores, 128GB of RAM, and two 24Gb NVIDIA RTX 3090 GPUs. The classifiers used were implemented in Python, through the use of Tensorflow [1], an open-source software library for machine learning, and Keras [2], a powerful and easy-to-use open-source Python library for developing and evaluating deep learning models.

The classifiers were trained by optimizing the hyperparameters to find the best possible configuration. More specifically, the hyperparameters tuned are reported in Table II, where in the second column there are the selected ranges adopted. Respectively, first (LR) represents the step size at each iteration as it moves towards a minimum loss function. The optimizer allows identifying, through a series of iterations, those weight values such that the cost function has the minimum value. The dropout rate (D0) is the fraction of neurons to be reset in a given level at each training phase. The deep supervision (DS) indicates precisely whether the internal nodes of the network are also included in the loss or if only the output has been considered [15]. In the first case, this parameter has the value yes, otherwise it does not. The batch size (BS) indicates the number of samples processed before the model was updated. Its size must be greater than or equal to one and less than or equal to the number of samples in the training dataset. Finally, batch normalization manages the normalization of the level inputs by re-centering and resizing and is used in 3 configurations: H, if placed at the beginning of the block, T at the end, N if not used.

[1]https://www.tensorflow.org/

[2]https://keras.io/

C. Validation

To validate the model we have adopted a metric that is usually used to evaluate segmentation: the sørensen-Dice similarity coefficient, which measures the similarity between two samples and is based on presence and absence data [7], and is similar to the Intersection over Union (IoU) metric, a good metric for measuring the overlap between two bounding boxes or masks [29].

Defined G as the correct mask and P as the segmentation generated by the network, the Dice can be defined as:

$$\text{Dice} = \frac{2 * |G \cap P|}{|G| + |P|}$$

To validate pixel classification performance and correctness of the segmentation area we adopted the sensitivity (SEN) and the positive predictive value (PPV) metrics, which are defined as:

$$\text{SEN} = \frac{|G \cap P|}{|G|}$$

$$\text{PPV} = \frac{|G \cap P|}{|P|}$$

VI. Results and Discussion

The table I summarizes the results obtained for GUNET3++ compared to the considered baseline methods (i.e., UNET, UNET++ and UNET3++). The training process uses a hyperparameter optimization step (HPO) on the intervals shown in the II table, so the table shows the best five hyperparameter configurations. To demonstrate the effectiveness of the GUNET3++ network, we too applied the same hyperparameter optimization to all basic methods to select the best-optimized network of each type. Above, in the table, for each type of network, we report the number of parameters used, in the first 5 columns we report the hyperparameters used, except for the UNET network, for which there are four columns because in-depth supervision is not used as it is not applicable. The last three columns show the performance of the proposed approach, evaluated using the above

UNET 37.39M params								UNET++ 49.39M params								UNET3+ 26.39M params								GUNET3++ 61.82M params							
DO	BN	LR	Opt	BS	Dice	SEN	PPV	DO	BN	LR	Opt	DS	BS	Dice	SEN	PPV	DO	BN	LR	Opt	DS	BS	Dice	SEN	PPV	DO	BN	LR	Opt	DS	BS	Dice	SEN	PPV
0.20	H	0.03	NADAM	32	0.728	0.807	0.708	0.30	H	0.04	SGD	Y	64	0.922	0.929	0.907	0.15	H	0.01	SGD	Y	64	0.934	0.936	0.918	0.25	H	0.03	SGD	Y	64	0.961	0.966	0.933
0.20	H	0.03	SGD	64	0.724	0.800	0.677	0.25	H	0.02	SGD	Y	32	0.918	0.928	0.895	0.10	H	0.01	SGD	Y	64	0.932	0.937	0.897	0.20	H	0.04	NADAM	Y	64	0.954	0.961	0.914
0.20	T	0.02	NADAM	64	0.696	0.768	0.661	0.20	T	0.03	SGD	N	64	0.912	0.926	0.863	0.10	T	0.02	NADAM	Y	32	0.924	0.926	0.878	0.20	T	0.02	NADAM	Y	64	0.944	0.950	0.902
0.15	H	0.02	NADAM	64	0.665	0.738	0.611	0.30	H	0.02	NADAM	N	32	0.891	0.913	0.837	0.15	H	0.02	NADAM	N	32	0.902	0.908	0.870	0.30	H	0.02	NADAM	N	128	0.917	0.924	0.835
0.25	N	0.01	RMSProp	32	0.661	0.733	0.605	0.25	T	0.02	SGD	Y	128	0.877	0.885	0.819	0.20	T	0.02	RMSProp	N	128	0.893	0.897	0.865	0.25	T	0.05	SGD	N	64	0.914	0.919	0.904

TABLE I: Results comparison of GUNET3++ with baseline methods (best five configuration of hyper-parameters).

Hyperparameters	Ranges
Learning rate	[0.01, 0.05]
Optimization algorithm	SGD, Adam , RmsProp, Nadam
Dropout rate	[0.10, 0.25]
Deep supervision	Yes, No
Batch size	{ 16, 32, 64, 128, 256}
Batch Normalization	Head, Tail, None

TABLE II: Hyper-parameters Optimization and selected ranges.

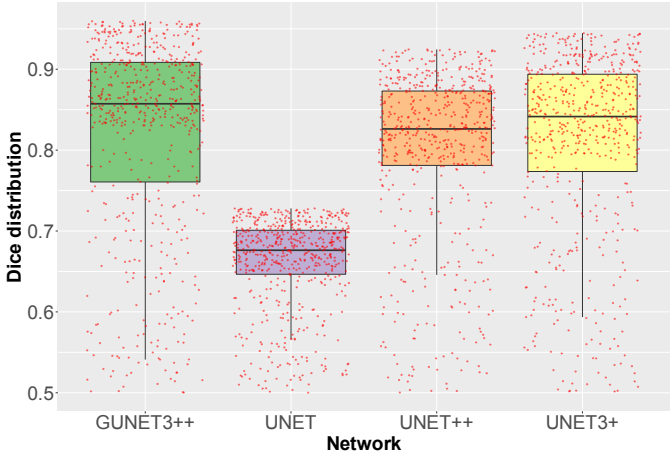


Fig. 4: Boxplot of dice score distribution of hyper-parameter optimization process for all fine-tuned networks.

metrics (Dice, SEN, PPV). Finally, we highlight the best configuration in yellow. The results obtained show a significantly improved value for the average Dice by $\approx 3\%$. The reason behind such improvement is due to the ability of GUNET3++ of using contextual information at the same resolution using horizontal skip connections jointly with multi-scale operations. This help to reduce the size of the predicted ROI allowing the network to better focus on the nodule effectively ignoring redundant information (i.e., other kinds of tissues that are not relevant).

Figure 4 shows the boxplots of the Dice score of each HPO process execution for each of the fine-tuned networks. As we can see standard UNET is the worst network in terms of Dice average (all configurations have a Dice lower than ≈ 0.72). It is also interesting that with respect to other networks, the number of bad configurations is significantly higher meaning that the training process is less stable. The other two networks (UNET++ and UNET3+) provide much better results with several configurations reaching a Dice of ≈ 0.92 for UNET++ and ≈ 0.93 for UNET3+. As the figure shows, GUNET3++ sits in the middle for training stability, having a lower number of unsuccessful configurations than UNET but higher concerning UNET++ and UNET3+. This is due to the number of parameters that are substantially higher in GUNET3++ with respect to other networks, making the training process more difficult and less stable. Conversely, GUNET3++ is the best performing network in terms of Dice having a high number of configurations reaching an average Dice equals to ≈ 0.96.

This is also noticeable by the qualitative analysis shown in Fig.5 where the reference annotations agreed by expert radiologists are compared with ROI extracted by predicted masks. As the figure shows, the other networks are less capable of precisely identifying the nodule borders leading to a worse Dice concerning GUNET3++. This is especially true for more complex regions where the morphology of the nodule's border is irregular and unsymmetrical. This is confirmed, qualitatively, by the example in the second row: the asymmetric nodule is better identified with respect to the other networks (UNET++ is unable to identify the irregular boundary).

From a quantitative point of view, this is also confirmed by the distribution analysis of Dice scores for best-fine-tuned networks with respect to all segmented nodules. The histogram in Fig. 6 shows the distribution of Dice for the LIDC-IDRI test set. As we can see, the figure highlights that most of the test instances for both UNET++ and UNET3+ are above 0.85, indicating good performance for all those networks with respect to standard UNET that is more concentrated in the range $[0.75, 0.85]$. It is interesting however to observe that with respect to both UNET++ and UNET3+, GUNET3++ Dice values are more concentrated in the higher narrower interval $[0.85, 0.95]$, indicating better average performance.

VII. Threats to the validity

In this study, it is possible to identify three different types of threats to validity: construct, internal and external. Respectively, for the first, the dataset adopted for the experiment could represent a possible limitation. To overcome this problem, we chose a dataset that has been

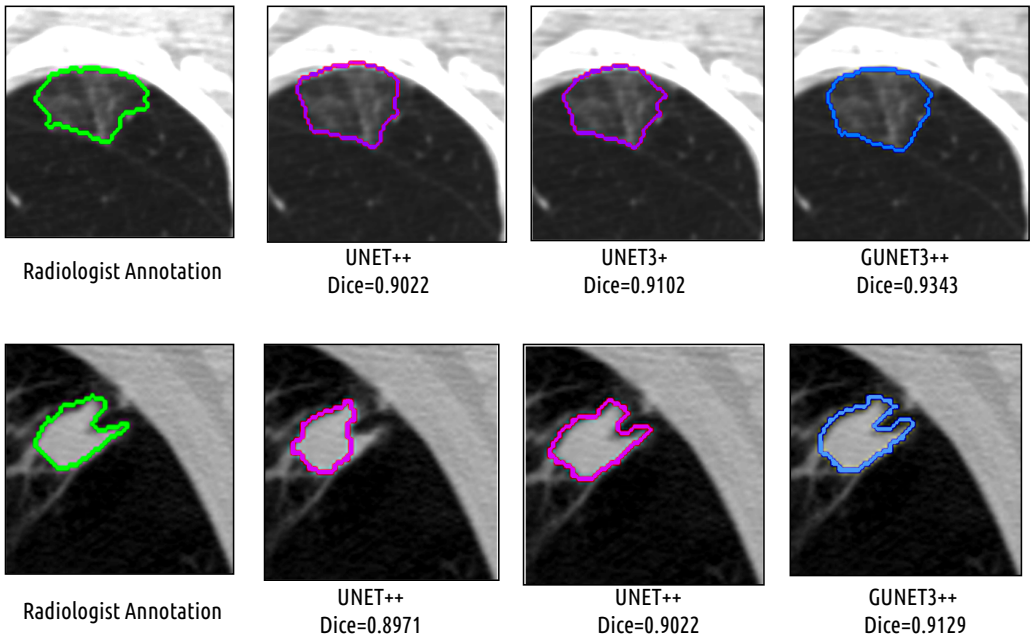


Fig. 5: Qualitative analysis of predicted ROI.

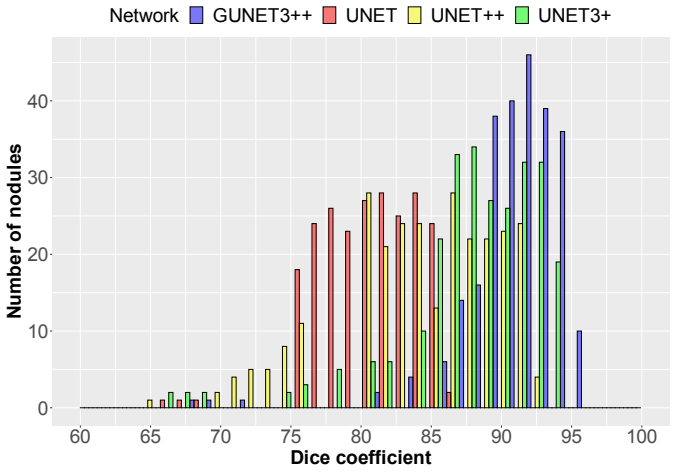


Fig. 6: Best fine-tuned networks histogram of Dice distribution over test-set.

referenced in numerous medical and engineering studies. Furthermore, a pre-processing and cleaning phase was carried out to filter all images of different formats which had low quality. Finally, a data labeling process was carried out which involved four radiologists. In addition, some patients were not considered because there were parsing errors in the metadata, or there were overlapping errors of different acquisitions. For internal validity threats, the results could be influenced by variables not considered. The dataset has been split into training and test sets with a percentage of 80/20, and we can't know if different splits would lead to the same conclusions or not. Furthermore, the performance of the approach is affected by the network architectures we have chosen, and different architectures may lead to a variation in the results. To mitigate this threat, we evaluated four different network configurations in the study. Finally, as far as threats to external validity are concerned, these concern the generalization of results. The data set considered includes a large number of patients and related images, but it would be necessary to test the approach on datasets containing images with different formats, resolutions, and colors.

VIII. Conclusions

The problem of lung cancer detection has been addressed in many recent studies. The centrality of the problem stems from the importance of early lung cancer detection so that patients have a longer life expectancy. In this work, a novel approach has been proposed whose goal is to detect the presence of lung cancer using CT scan images, which are used to train an enhanced version of the UNET segmentation network. The proposed approach was validated using the LIDC-IDRI dataset.

The proposed network, GUNET3++, is the best performing network, compared to the other variants of UNet, in terms of Dice having a high number of configurations reaching an average Dice of about 0.96. Furthermore, even considering a qualitative analysis, GUNET3++ is more able to precisely identify the boundaries of the nodule leading to a better Dice than the other variants. A possible future development could be the expansion of the dataset used to generalize the results.

REFERENCES

[1] Pasquale Ardimento, Lerina Aversano, Mario Luca Bernardi, and Marta Cimitile. Deep neural networks ensemble for lung nodule detection on chest ct scans. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2021.

[2] Lerina Aversano, Mario Luca Bernardi, Marta Cimitile, Martina Iammarino, Paolo Emidio Macchia, Immacolata Cristina Nettore, and Chiara Verdone. Thyroid disease treatment prediction with machine learning approaches. In Jaroslaw Watróbski, Wojciech Salabun, Carlos Toro, Cecilia Zanni-Merk, Robert J. Howlett, and Lakhmi C. Jain, editors, *Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 25th International Conference KES-2021, Virtual Event / Szczecin, Poland, 8-10 September 2021*, volume 192 of *Procedia Computer Science*, pages 1031–1040. Elsevier, 2021.

[3] Lerina Aversano, Mario Luca Bernardi, Marta Cimitile, and Riccardo Pecori. Fuzzy neural networks to detect parkinson disease. In *29th IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2020, Glasgow, UK, July 19-24, 2020*, pages 1–8. IEEE, 2020.

[4] Lerina Aversano, Mario Luca Bernardi, Marta Cimitile, and Riccardo Pecori. Deep neural networks ensemble to detect COVID-19 from CT scans. *Pattern Recognit.*, 120:108135, 2021.

[5] Ujjwal Baid, Sanjay Talbar, Swapnil Rane, Sudeep Gupta, Meenakshi H. Thakur, Aliasgar Moiyadi, Nilesh Sable, Mayuresh Akolkar, and Abhishek Mahajan. A novel approach for fully automatic intra-tumor segmentation with 3d u-net architecture for gliomas. *Frontiers in Computational Neuroscience*, 14:10, 2020.

[6] Maria G. Baldeon Calisto and S. Lai-Yuen. Adaresu-net: Multiobjective adaptive convolutional neural network for medical image segmentation. *Neurocomputing*, 392:325–340, 2020.

[7] Aaron Carass, Snehashis Roy, Adrian Gherman, Jacob C. Reinhold, Andrew Jesson, Tal Arbel, Oskar Maier, Heinz Handels, Mohsen Ghafoorian, Bram Platel, Ariel Birenbaum, Hayit Greenspan, Dzung L. Pham, Ciprian M. Crainiceanu, Peter A. Calabresi, Jerry L. Prince, William R. Gray Roncal, Russell T. Shinohara, and Ipek Oguz. Evaluating white matter lesion segmentations with refined sørensen-dice analysis. *Scientific reports*, 10(1):8242–8242, May 2020. 32427874[pmid].

[8] Jason L. Causey, Yuanfang Guan, Wei Dong, Karl Walker, Jake A. Qualls, Fred Prior, and Xiuzhen Huang. Lung cancer screening with low-dose ct scans using a deep learning approach, 2019.

[9] S. Das and S. Majumder. Lung cancer detection using deep learning network: A comparative analysis. In *2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, pages 30–35, 2020.

[10] S. Durga, R. Nag, and E. Daniel. Survey on machine learning and deep learning algorithms used in internet of things (iot) healthcare. In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, pages 1018–1022, 2019.

[11] Junfeng Gao, QingZeng Song, Lei Zhao, XingKe Luo, and XueChen Dou. Using deep learning for classification of lung nodules on computed tomography images. *Journal of Healthcare Engineering*, 2017:8314740, 2017.

[12] Ross Gruetzemacher, Ashish Gupta, and D. Paradice. 3d deep learning for detecting pulmonary nodules in ct scans. *Journal of the American Medical Informatics Association*, 25:1301–1310, 2018.

[13] Muzzamil Javaid, Moazzam Javid, Muhammad Zia Ur Rehman, and Syed Irtiza Ali Shah. A novel approach to cad system for the detection of lung nodules in ct images. *Computer Methods and Programs in Biomedicine*, 135:125–139, 2016.

[14] Guixia Kang, Kui Liu, Beibei Hou, and Ningbo Zhang. 3d multi-view convolutional neural networks for lung nodule classification. *PLOS ONE*, 12(11):1–21, 11 2017.

[15] Chi Li, M. Zeeshan Zia, Quoc-Huy Tran, Xiang Yu, Gregory D. Hager, and Manmohan Chandraker. Deep supervision with intermediate concepts, 2018.

[16] Ingo Lütkebohle. LIDC-IDRI - The Cancer Image Archive (TCIA) Public Access. https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI/, 2014. [Online; accessed 30-January-2021].

[17] Nasrullah Nasrullah, Jun Sang, Mohammad S. Alam, Muhammad Mateen, Bin Cai, and Haibo Hu. Automated lung nodule detection and classification using deep learning combined with multiple strategies. *Sensors*, 19(17), 2019.

[18] International Agency For Research on Cancer. Estimated age-standardized mortality rates (World) in 2020, worldwide, both sexes, all ages. https://gco.iarc.fr/today/data/factsheets/cancers/39-All-cancers-fact-sheet.pdf, 2020. [Online; accessed 01-February-2021].

[19] International Agency For Research on Cancer. Estimated number of new cases from 2020 to 2040, Both sexes, age [0-85+] Trachea, bronchus and lung. https://gco.iarc.fr/tomorrow/en/dataviz/bars?mode=population&cancers=15, 2020. [Online; accessed 01-February-2021].

[20] World Health Organization. Knowledge into Action Cancer Control - WHO Guide for Effective Programmes. https://www.who.int/cancer/modules/Early%20Detection%20Module%203.pdf, 2020. [Online; accessed 04-February-2021].

[21] Narinder Singh Punn and Sonali Agarwal. Inception u-net architecture for semantic segmentation to identify nuclei in microscopy cell images. *ACM Trans. Multimedia Comput. Commun. Appl.*, 16(1), February 2020.

[22] Diego Riquelme and Moulay A. Akhloufi. Deep learning for lung cancer nodules detection and classification in ct scans. *AI*, 1(1):28–67, 2020.

[23] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, 2015.

[24] Evan Shelhamer, Jonathan Long, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651, 2017.

[25] Wei Shen, Mu Zhou, Feng Yang, Dongdong Yu, Di Dong, Caiyun Yang, Yali Zang, and Jie Tian. Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. *Pattern Recognition*, 61:663–673, 2017.

[26] Run Su, Deyun Zhang, Jinhuai Liu, and Chuandong Cheng. Msu-net: Multi-scale u-net for 2d medical image segmentation. *Frontiers in Genetics*, 12:140, 2021.

[27] Wenqing Sun, Bin Zheng, and Wei Qian. Automatic feature learning using multichannel roi based on deep structured algorithms for computerized lung cancer diagnosis. *Computers in Biology and Medicine*, 89:530–539, 2017.

[28] Nima Tajbakhsh and Kenji Suzuki. Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification. *Pattern Recogn.*, 63(C):476–486, March 2017.

[29] Floris van Beers, Arvid Lindström, Emmanuel Okafor, and Marco A Wiering. Deep neural networks with intersection over union loss for binary image segmentation. In *ICPRAM*, 2019.

[30] Michael W. Vannier, Ayman El-Baz, Garth M. Beache, Georgy Gimel'farb, Kenji Suzuki, Kazunori Okada, Ahmed Elnakib, Ahmed Soliman, and Behnoush Abdollahi. Computer-aided diagnosis systems for lung cancer: Challenges and methodologies. *International Journal of Biomedical Imaging*, 2013:942353, 2013.

[31] C. White, T. Flukinger, J. Jeudy, and Joseph J Chen. Use of a computer-aided detection system to detect missed lung cancer at chest radiography. *Radiology*, 252 1:273–81, 2009.

[32] Kun-Hsing Yu, Tsung-Lu Michael Lee, Ming-Hsuan Yen, S C Kou, Bruce Rosen, Jung-Hsien Chiang, and Isaac S Kohane. Reproducible machine learning methods for lung cancer detection using computed tomography images: Algorithm development and validation. *J Med Internet Res*, 22(8):e16709, Aug 2020.

[33] Guobin Zhang, Zhiyong Yang, Li Gong, Shan Jiang, Lu Wang, Xi Cao, Lin Wei, Hongyun Zhang, and Ziqi Liu. An appraisal of nodule diagnosis for lung cancer in ct images. *Journal of Medical Systems*, 43(7):181, 2019.

[34] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, 39(6):1856–1867, Dec 2019.

[35] W. Zhu, C. Liu, W. Fan, and X. Xie. Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 673–681, 2018.