



Complete fully automatic detection, segmentation and 3D reconstruction of tumor volume for non-small cell lung cancer using YOLOv4 and region-based active contour model

Sifundvolesihle Dlamini ^a, Yi-Hsi Chen ^b, Chung-Feng Jeffrey Kuo ^{a,*}

^a Department of Materials Science and Engineering, National Taiwan University of Science and Technology, Taipei 106, Taiwan, ROC

^b Division of Thoracic Surgery, Department of Surgery, Tri-Service General Hospital, Keelung Branch, National Defense Medical Center, 100, Zhengrong St. Zhongzheng Dis., Keelung City 202, Taiwan, ROC



ARTICLE INFO

Keywords:
Lung tumor
Detection
Segmentation
Active contour
YOLOv4

ABSTRACT

We aim to develop a fully automatic system that will detect, segment and accurately reconstruct non-small cell lung cancer tumors into space using YOLOv4 and region-based active contour model. The system consists of two main sections which are detection and volumetric rendering. The detection section is composed of image enhancement, augmentation, labeling and localization while the volumetric rendering is mainly image filtering, tumor extraction, region-based active contour and 3D reconstruction. In this method the images are enhanced to eliminate noise before augmentation which is intended to multiply and diversify the image data. Labeling was then carried out in order to create a solid learning foundation for the localization model. Images with localized tumors were passed through smoothing filters and then clustered to extract tumor masks. Lastly contour information was obtained to render the volumetric tumor. The designed system displays a strong detection performance with a precision of 96.57%, sensitivity and F1 score of 97.02% and 96.79% respectively at a detection speed of 34 fps, prediction time per image of 21.38 ms. The system segmentation validation achieved a dice score coefficient of 92.19 % on tumor extraction. A 99.74 % accuracy was obtained during the verification of the method's volumetric rendering using a 3D printed image of the rendered tumor. The rendering of the volumetric tumor was obtained at an average time of 11 s. This system shows a strong performance and reliability due to its ability to detect, segment and reconstruct a volumetric tumor into space with high confidence.

1. Introduction

Cancer is one of the lethal diseases ever known to mankind accounting for nearly 10 million deaths every year. It is caused by the mutation of normal cells into tumors in a multi-stage process that progresses from a pre-cancerous lesion to a malignant tumor as a result of the interaction of a person's hereditary elements (World Health Organization, 2021). Lung cancer is the most common cause of cancer mortality and morbidity, accounting for one out of every-five cancer deaths. Its death rate exceeds that of the next two common cancers (colon and liver) combined (Siegel et al., 2019). A lung tumor begins in the tissues of the lungs, usually in the pulmonary cells lining the air passages and proliferates into multiple cells. The two common types of tumors are small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) according to their appearance under the microscope (Nature

Portfolio, 2021). NSCLC accounts for around 80 % of all lung tumors, with about 30 % being identified to be in advanced stage on their first examination. The majority of NSCLC patients who are diagnosed with early stage cancer tend to have surgery while up to more than half of them develop local and distant metastases (Novello & Le Chevalier, 2003).

Studies have shown that early precise detection and automatic tumor volumetry can not only tackle the issue of metastases, but can also assist in surgical procedures and the assessment of treatment response, and save a vast number of lives (Godoy et al., 2013; Jacobson & Jaklitsch, 2018; Neal Joshua et al., 2021). It has been proven that automated volumetry provides superior reproducibility response evaluations and should be favored over alternative measurement methods (Marten et al., 2006; Mathew et al., 2020). Using regular shaped phantoms, Sohaib et al. (2000) demonstrated that computed tomography (CT) volume

* Corresponding author.

E-mail address: jeffreykuo@mail.ntust.edu.tw (C.-F. Jeffrey Kuo).

measurement has less than 5 % error, and that there is a 90 % agreement between one-dimensional (1D) and two-dimensional (2D) measurements in the assessment of treatment, and a 100 % agreement between 2D and three-dimensional (3D) assessments. [Orsatti et al. \(2020\)](#) also proved that 3D radiological assessment predicts survival better than 2D and 1D measurement in malignant cells. A study was conducted to improve the response evaluation criteria in solid tumor (RECIST) evaluation criteria for NSCLC patients by employing a 3D measurement of tumor volume, and it was discovered that it has good reconstruction and high accuracy ([Mozley et al., 2010](#)). When compared to 1D and 2D diameter measures, 3D based metrics have a stronger correlation with clinical findings. The evidence clearly indicates that volumetric image analysis adds significant value to clinical treatment, however, a more precise automatic tumor volumetric system is yet to be developed.

The development of an automated tumor volumetric systems has recently attracted a lot of attention, and several algorithms and techniques have been put forward to deal with the latter. Clustering plays a crucial role in segmentation through the partition of images into a number of disjoint clusters with similar properties based on color and texture features. Thorax CT images display the same tissues with similar features, which makes the samples in our study suitable for clustering. *k*-means is one of the simplest and widely used unsupervised machine-learning clustering algorithm that makes inference from datasets using input vectors without referencing to known results. [Ng et al. \(2006\)](#) proposed a *k*-means clustering that incorporates the watershed segmentation algorithm for medical image analysis. Since watershed is prone to over segmentation and sensitive to false edges, *k*-means clustering was used to produce primary segments prior to the application of the watershed algorithm. [Shrivastava et al. \(2014\)](#) manipulated *k*-means to achieve a high performing algorithm to cluster brain tumor magnetic resonance images (MRI). [Katkar et al. \(2015\)](#) used *k*-means with principle component analysis (PCA) for feature extraction and formation of precise number of clusters to increase accuracy in the segmentation of MRI. Subtractive clustering was utilized by [Dhanachandra et al. \(2015\)](#) to generate centroids based on the potential value of the data points before applying *k*-means. The generated centers were then used with the *k*-means algorithm for the segmentation of images. Our study adopted *k*-means clustering algorithm due to its nonlinearly separation capability, speed and simplicity ([Ghosh & Dubey, 2013](#)).

Contour extraction is crucial for the accurate reconstruction of tumor tissues in 3D. There are plenty of techniques for segmenting the region of interest that can be leveraged. [Osher and Sethian \(1988\)](#) presented non-oscillatory algorithm schemes to accurately capture sharp gradients in moving fronts. These schemes aimed at solving Hamilton-Jacobi equations with parabolic right-hand sides by using the approximation techniques from hyperbolic conservation laws. They were designed to handle topological merging in spatial dimensions. [Li et al. \(2010\)](#) proposed a distance regularized level set evolution that improves contour extraction accuracy by correcting re-initialization of level set approaches. However, this method is incapable of dealing with blurred images since it uses image gradient as a condition of termination. Active contour segmentation is one of the most common techniques used in medical images. [Hemalatha et al. \(2018\)](#) presented a few of the most successful active contour algorithms in medical image analysis. [Chan and Vese \(2001\)](#) came up with an active contour model based on the technique of curve, Mumford-Shah function and level set for segmenting objects in space. This model uses energy minimization to segment objects with boundaries that are not determined by gradient. Region-based active contour introduced by [Li et al. \(2008\)](#) was embraced by our study to segment and extract contour information. Our study took inspiration from this technique because of its strong capability of segmenting using intensity information in local region at a controllable scale. A successful and robust implementation of region-based active contour was executed by [Ma et al. \(2019\)](#) to segment synthetic images and real medical images.

Nowadays, the most popular and more reliable way of dealing with

medical image analysis is using deep learning techniques. With this approach, classifiers are trained to distinguish between tumorous and non-tumorous nodules using labeled data. Convolutional networks have already proven that they can handle complex data and have a strong error tolerance, parallel computing, generalization and self-learning capabilities. In one study, a neural network was employed to detect lung cancer using physical symptoms such as yellow fingers, anxiety, allergy, coughing, chest pain, et cetera ([Nasser & Abu-Naser, 2019](#)). [Lakshmanaprabu et al. \(2019\)](#) presented an innovative approach to diagnose lung cancer on CT images using Optimal Deep Neural Network and Linear Discriminant Analysis. The neural network is applied on the images and optimized using a Modified Gravitational Algorithm for cancer identification. A deep Convolutional Neural Network (CNN) was used by [Zhang et al. \(2019\)](#) to integrate a well-trained model to detect and classify tumor nodules. Competitive results were also achieved by [Kasinathan et al. \(2019\)](#) through the combination of CNN and active contours to detect and classify 3D lesion lung tumor CT images automatically.

YOLO convolutional neural network has already been heavily applied in a variety of industries to detect object of interests in real-time ([Cao et al., 2021; Zhang et al., 2021; Zhu & Spachos, 2021](#)). [George et al. \(2018\)](#) successfully applied YOLO as well in medical images to localize tumor nodules from low dose CT images and obtained good results. YOLO is a deep convolutional neural network that detects objects in real-time by predicting bounding boxes and class probabilities. It is part of the one-stage detector family often recognized as one shot detectors ([Redmon, 2016](#)). Since YOLO's performance is generally faster than most of modern real-time object detection models, this study opted to adopt YOLOv4 to solve the detection speed and precision issues in the localization of tumor nodules. What makes YOLO a model of preference in our study is the fact that it makes the composition process easier and more effective due to the use of an integration detection framework which includes candidate frame extraction, convolutional learning features and non-maximum suppression optimization ([Neubeck & Van Gool, 2006](#)). More optimization add-ons were attached to the basic YOLO network structure such as Batch Norm ([Ioffe & Szegedy, 2015](#)), anchor box convolutions ([Chen et al., 2019](#)) and dimension cluster ([Bouveyron et al., 2007](#)) to upgrade the model to a YOLO9000 ([Redmon & Farhadi, 2017](#)). The merging of the residual network with feature pyramid network (FPN) ([Lin et al., 2017](#)) using binary loss function gave birth to a YOLOv3 ([Redmon & Farhadi, 2018](#)) model. This model make use of YOLO9000, ResNet ([He et al., 2016](#)) and DarkNet with Batch Norm and DropOut ([Srivastava et al., 2014](#)) operations added at each stage of convolution to extract features. YOLOv4 ([Bochkovskiy et al., 2020](#)) was built around two innovative techniques; the bag of freebies (BoF) and bag of specials (BoS). BoF includes pre-processing techniques and extra modules such as CutMix and Mosaic augmentation ([Yun et al., 2019](#)), and DropBlock ([Ghiasi, 2018](#)) respectively which ensures inference enhancement without computational increase, while the BoS is plugin modules and post-processing techniques to ameliorate the accuracy. Unique add-ons in YOLOv4 includes SSP-block ([He, 2015](#)) and PAN-block ([Liu et al., 2018](#)) and a YOLOv3 as a Dense prediction module.

Although deep learning techniques are very reliable with good self-learning capabilities and overall generalization, there's still no good enough evidence of real-time detection with satisfactory precision in lung tumors. This is due to that a lot of models still relies on proposal generator that generates poor and sparse sets of proposals and extract features from each followed by region classifiers that predicts the category of proposed region. Furthermore, contours are still extracted using distance regularized level set evolution which have a higher contour extraction accuracy but incapable of dealing with blurred image boundary detection due to their use of image gradient as a stop condition for curve evolution. These shortcomings inspired our interest to develop a relatively better system to detect, segment and reconstruct lung tumors.

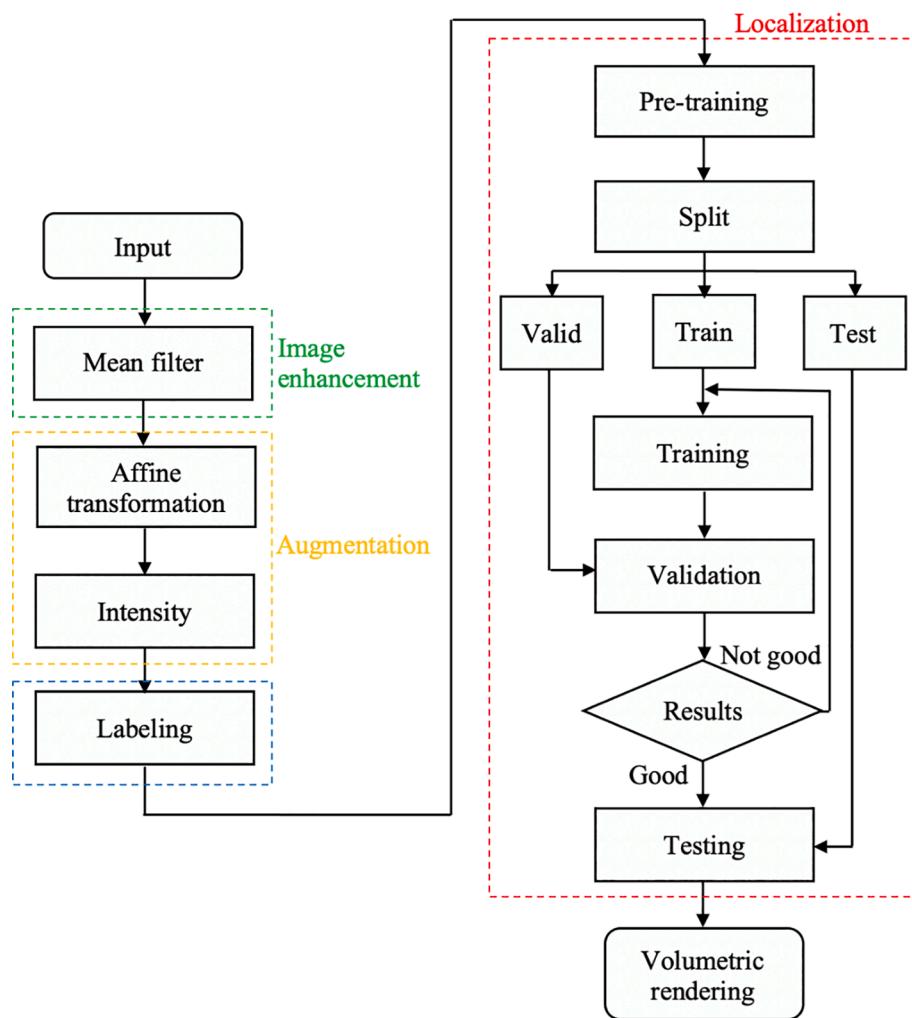


Fig. 1. Detection schematics; image enhancement, augmentation, labeling, localization.

The design of our study aimed to develop a fully automatic system to detect, segment and reconstruct lung tumor volumetry. The system took advantage of YOLOv4 model to achieve good detection speed and localization precision. The localization model was trained from our pre-trained weights trained from scratch using augmented images passed through a filter to eliminate CT noise. k-means was employed to partition images into disjoined clusters to obtain tumor masks. To extract tumor contour information, region-based active contour was used followed by marching cube algorithm to reconstruct the 3D volumetric tumor.

The main contributions of this study are: 1) An enhanced tumor detection algorithm incorporating numerous data preprocessing techniques with YOLOv4 model to achieve good detection precision in real-time. 2) An advanced tumor rendering algorithm utilizing a non-image gradient dependent active contour model customized for image series contour extraction with an automation feature.

The study is presented as follows; the second section is the description of the methodology, third section is the detailed explanation of the experiment and results, while fourth and fifth is the system validation

and discussion respectively, and lastly is the conclusion of the study.

2. Methodology

We propose a system that will detect, segment and reconstruct NSCLC tumor volumetry automatically using YOLOv4 and region-based active contour. To design the system, we used a dataset for patients with NSCLC who underwent surgery, acquired from the cancer imaging archive (TCIA) (Clark et al., 2013) which is a free access platform for medical images. This data was compiled and published by Quantitative Imaging Network (QIN) (2021) with the goal of improving the role of quantitative imaging in cancer clinical decision making. The lung image database consortium image collection and image database resource initiative (LIDC-IDRI) (Armato et al., 2011) dataset from TCIA as well was used to test the proposed system functionality. The system proposed in this study is mainly divided into 2 sections; tumor detection and volumetric rendering.

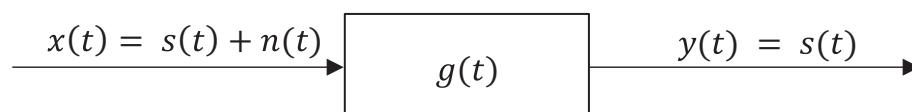


Fig. 2. Mean filter system diagram.

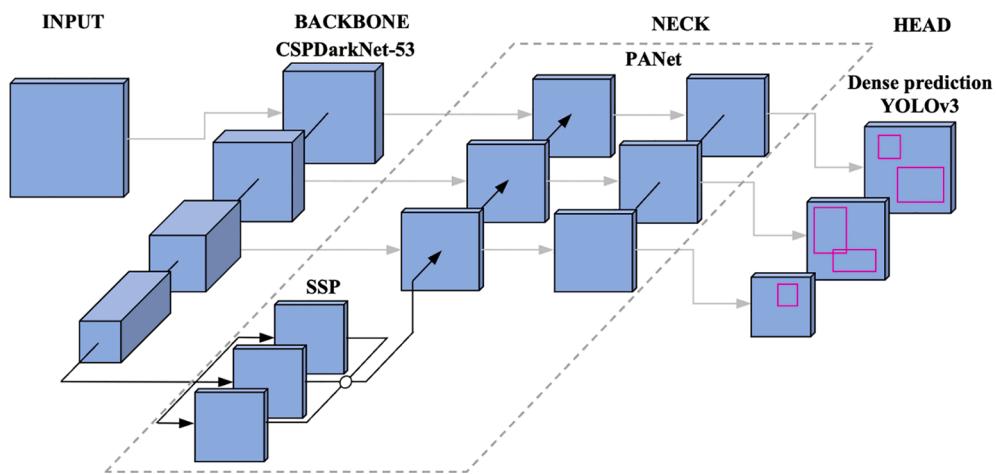


Fig. 3. YOLOv4 architecture.

2.1. Tumor detection

The aim in this section is to automatically detect and locate tumor nodules as regions of interest. This section has four parts, which are image enhancement, augmentation, labeling, and localization as shown in Fig. 1.

2.1.1. Image enhancement

Mean filter was used as an enhancement tool to reduce noise in the images (Gupta, 2011). This filter is a low pass, linear filter which eliminates noise by reducing the amount of intensity of variation between pixels. In the typical Mean filtering technique, the value of each pixel is replaced with the mean of its neighbors including itself. Fig. 2 shows a flow diagram explaining the Mean filter, with $g(t)$ being the filter while $x(t)$ representing the input signal containing the real signal $s(t)$ and noise $n(t)$, and $y(t)$ is the output signal as expressed in equation (1).

$$y(t) = g(t)^*(s(t) + n(t)) \quad (1)$$

For real images, let S_{xy} represent the set of coordinates in a rectangle sub image window of size $m \times n$ centered at point (x, y) . In the area specified by S_{xy} , the Mean Filter kernel computes the average value of the noisy image $g(s, t)$. The equation of computation is shown in equation (2) below.

$$f(x, y) = \frac{1}{mn} \sum_{(s, t) \in S_{xy}} g(s, t) \quad (2)$$

This process is implemented using a convolutional kernel which implies that all of the coefficients have the same value of $1/mn$. The elimination of the noise is due to the process of smoothing of local variation (Sun, 1994).

2.1.2. Augmentation

The images were augmented after they were enhanced. Augmentation is the technique of creating data variation while multiplying the data to be able to develop a flexible and more robust model. We employed affine transformation in our augmentation such as 90° , 180° & 270° rotation, as well as horizontal and vertical flips. In addition, the light intensity of each of the altered images was increased and decreased.

2.1.3. Labeling

Tumors on the data sample were manually located and tagged by experienced personnel in the process of image labeling. This is a very important step in supervised learning because the quality of the training data has a significant impact on the quality of the detection model. We

utilized OpenLabeling (Cartucho et al., 2018), a freely available graphical labeling source codes based on Python. A user interface is prompted by the codes where the tumor nodules are manually labeled and bounding boxes are drawn around them. This process is done to provide the localization model a solid learning foundation. The labeled images are then stored in a *.txt file format which the localization model can read.

2.1.4. Localization

The goal of localization is to pinpoint the tumor nodules. This study opted to use YOLOv4 as a localization model in the proposed system (Bochkovskiy et al., 2020). This model does not only solve the detection issues but it also deals with the regression issues experienced by a lot of detection models in our days. The three key attributions for the detection of tumors using this model are: precise detection prediction, good classification prediction and prediction across various scales.

YOLOv4 detectors are made up of three primary building blocks which are the backbone, neck and the head as shown in Fig. 3. The input images are processed by the convolutional backbone, which generates a high-level representation of the nodules based on their features. This representation is down-sampled to select deeper features that produce better prediction decision boundaries. At the detection step, many bounding boxes are created and mapped across the image with a nodule using aggregated convolutional backbone features in the detector's neck. The head execute the class prediction and location at the same time.

The backbone model used by YOLOv4 is cross stage partial network (CSPDarkNet-53) which is a modified type of a DenseNet. CSPDarkNet-53 addresses the problem of vanishing gradients, allowing the entire model to learn using feature propagation and re-usability. This backbone model also reduces the network's hyperparameters which assures shorter training latency and build up stronger decision boundaries. On top of the adopted backbone model, YOLOv4 is additionally concatenated with a spatial pyramid pooling (SPP) block to improve deep receptive fields without incurring computational complexity.

To improve discriminative ability of the region of interest representative, the tumor features are aggregated using path aggregation network (PANet) in the neck. Localization is enhanced by the use of BoF and BoS which are additional training optimization. During training, the BoF assures improvement without incurring inference costs, whereas the BoS are add-on postprocessing packages that greatly influence the model's performance. BoF is mainly data augmentation in which the images are distorted for better benefits with CutMix and Mosaic, and DropBlock regularization is used to boost the model's learning capabilities. YOLOv4 model is also created by combining self-adversarial training (SAT) with a genetic algorithm to choose hyperparameters.

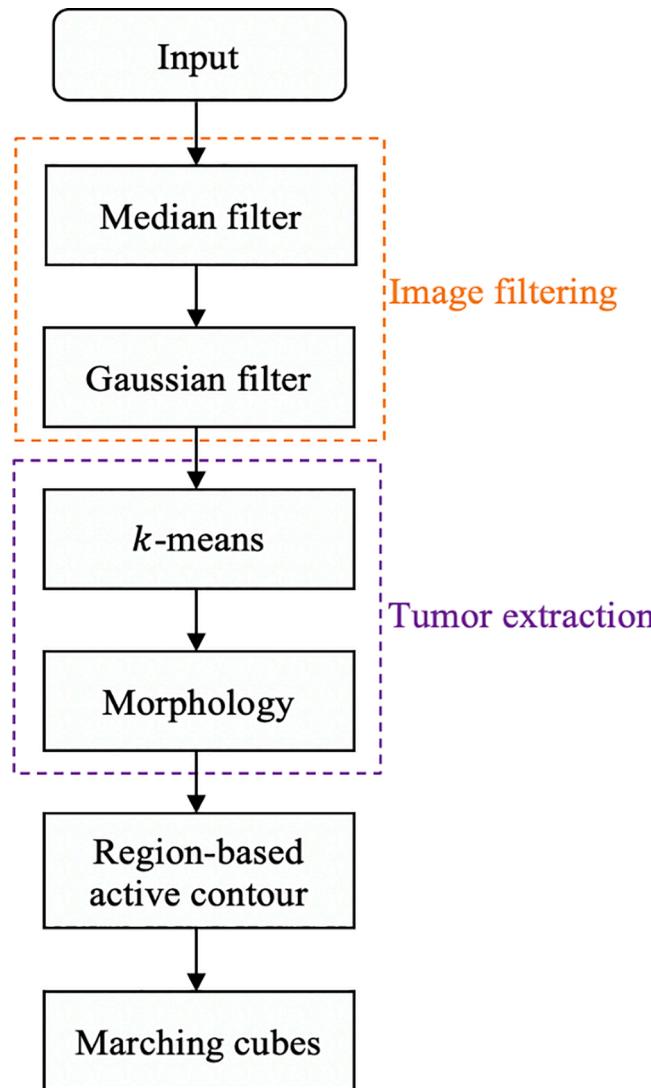


Fig. 4. Tumor volumetric rendering schematics; image filtering, tumor extraction, region-based active contour, marching cubes.

After the model has been thoroughly trained, test images are provided for nodules detection. A YOLOv3 dense prediction model is used to dissects the images into tiny grids using contextual information from the whole image to execute a regression operation to predict the position and identification of the tumor in a single pass. It then evaluates the characteristics of the test images and creates a tensor of the size $n \times n \times [(C+1)+\text{class}] \times k$ as an output, where $[(C+1)+\text{class}]$ is the bounding box offsets with the class predictions constant for the total number of classes respectively, multiplied everything by k the number of anchors. The constant C represents (x, y) coordinates, width, height and confident score which are the four bounding box components. The centroid of the bounding box in relation to the grid is represented by the (x, y) coordinates when the width and the height are calculated in terms of the entire image.

This model is an end-to-end training and a real-time processing with good results. When the ground truth T falls in the $n \times n$ grid, the tumor nodule is detected. Every grid is in charge of determining bounding boxes as well as confidence scores and class probability. The confidence score C_f in equation (3) is used determine whether or not the region of interest contains a nodule of interest.

$$C_f = p_m \times IoU_{pred}^T, \quad p_m \in \{0, 1\} \quad (3)$$

If the tumor nodule fell within the vicinity of the grid, the probability of occurrence of a tumor becomes 1 ($p_m = 1$), otherwise 0 ($p_m = 0$). The IoU_{pred}^T reflects the existence of a tumor in the ground truth T bounding box. C_f calculates the accuracy of the tumor detection. When the bounding boxes indicates the existence of a tumor in the same region, the model uses a baseline overlapping score to choose the optimum bounding box. Furthermore, the anticipated nodule inside the bounding boxes is weighted by estimated probabilities for the bounding box which is necessary for computing the confidence score and improve precision.

2.2. Tumor volumetric rendering

Image processing techniques were used with an aim to segment, calculate and render the volume of the nodules localized. The used techniques were image filtering, k -means, region-based active contour, and marching cubes respectively as shown in Fig. 4.

2.2.1. Image filtering

2.2.1.1. Median filter. The median filter is a great way to get rid of noise while preserving crucial features such as edges and lines (Zhu & Wang, 2012). This filter is a low pass non-linear filter that eliminates impulse noise by replacing each pixel's gray level with the median of the gray level in its neighborhood. The filtering is achieved by sliding a window over the image. A median filter is mathematically expressed as shown in equation (4).

$$y[m, n] = \text{median}\{x[i, j]; (i, j) \in \omega\} \quad (4)$$

where ω defines the neighborhood centered around the location $[m, n]$ in the image.

2.2.1.2. Gaussian filter. A gaussian filter is a non-linear filter used to blur images (Deng & Cahill, 1993). It is a low pass filter that removes noise using a Gaussian 'bell-shaped' kernel. The impulse response of a 1D gaussian filter is provided by equation (5) while the frequency response is given by equation (6) using f as the standard frequency. These equations can also be expressed with standard deviation as equation (7) and the frequency response is given by equation (8).

$$g(x) = \sqrt{\frac{a}{\pi}} e^{-ax^2} \quad (5)$$

$$h(f) = e^{-\frac{x^2}{a}} \quad (6)$$

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (7)$$

$$h(f) = e^{-\frac{f^2}{2\sigma^2}} \quad (8)$$

In 2D, the gaussian filter can be defined as equation (9).

$$g(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (9)$$

where x represents the distance in the horizontal axis from the origin, y is the distance in the vertical axis from the origin and σ is the standard deviation of the Gaussian distribution (Dogra & Bhalla, 2014).

During the process of filtering, the frequency coefficient is not abruptly cut, but a gradual cut off method is applied instead. The central pixels tend to have a larger weighting than those on the periphery due to the fact that this filter attenuates frequency components that are further away from the center. At the edge of the filter coefficients are close to zero.

2.2.2. k -means clustering

To remove the lung area and extract the tumor nodules from the CT,

we used k -means which can handle difficult extraction of complex texture in the lungs. k -means is a clustering unsupervised learning algorithm that clusters data points into groups based on their similarities. It tries to minimize the distance of points in a cluster with their centroids using a Euclidean distance. The objective function of k -means is expressed in equation (10).

$$J(V) = \sum_{i=1}^c \sum_{j=1}^n (\|x_i - \nu_j\|)^2 \quad (10)$$

where c and n are the number of pixels and the number of pixel clusters respectively and $\|x_i - \nu_j\|$ represents the Euclidean distance between pixel x_i and cluster center ν_j . The objective function can be achieved by iterations through the process shown below.

1. Select cluster centers c randomly.
2. Between each data point and cluster centers, compute the distance.
3. Assign the data point to the cluster center with the shortest distance between cluster centers among all the cluster.
4. Using equation (11), recalculate the new cluster.

$$\nu_j = \frac{\sum_{i=1}^c x_i c_i}{\sum_{i=1}^c c_i} \quad (11)$$

5. Recalculate the distance between each data point and new obtained cluster centers.
6. Stop if no data points are reassigned; otherwise start over from step 3.

2.2.3. Region-based active contour

The method used to extract contour information in this study is the region-based active contour. Active contours have been rapidly used in image segmentation achieving good results in subpixel accuracy of object boundaries. The aim of region-based model is to find the region of interest by guiding the active contour's motion with a specific region descriptor. The procedure is shown below.

1. Describe a fitting energy in terms of a contour and two fitting functions that locally approximate the image intensities on the two side of the contour.
2. Integrate the energy function into a variation level set formulation with a level set regularization term.
3. Generate a curve evolution equation for energy minimization.
4. Minimize the energy function to obtain objective function.

The approach is based on Chan and Vese (2001) aiming to minimize the Mumford-Shah (Mumford & Shah, 1989) function with an aim of defining a fitting energy. The Mumford-Shah energy function is defined as equation (12) when the image domain is $\Omega \subset k^2$ and $I : \Omega \rightarrow k$.

$$F^{MS}(u, C) = \int_{\Omega} (u - I)^2 dx + \mu \int_{\Omega/C} |\nabla u|^2 dx + \nu |C| \quad (12)$$

where $|C|$ is the length of the contour C , u is the approximation of the input image I , Ω is the image domain separated by C , μ and ν are weight coefficients.

Chan minimized the above equation (12) by using a piecewise constant function instead of image u as shown in equation (13).

$$F^{CV}(C, c_1, c_2) = \lambda_1 \int_{outside(C)} |I(x) - c_1|^2 dx + \lambda_2 \int_{inside(C)} |I(x) - c_2|^2 dx + \nu |C| \quad (13)$$

where $outside(C)$ and $inside(C)$ are the regions outside and inside C respectively, c_1 and c_2 are constants that estimates the intensities of image, λ_1 , λ_2 and ν are weight coefficients greater than 0.

Intensity information in discrete areas is utilized by the region-based

model at a controllable scale. Consider C be a closed contour in image domain Ω which separate Ω into two regions: $\Omega_1 = inside(C)$ and $\Omega_2 = outside(C)$, for a given point $x \in \Omega$, the local intensity fitting energy is mathematically expressed as equation (14).

$$\mathcal{E}_x^{Fit}(C, f_1(x), f_2(x)) = \sum_{i=1}^2 \lambda_i \int_{\Omega_i} K(x-y) |I(y) - f_i(x)|^2 dy \quad (14)$$

where λ_i are positive constants and $f_i(x)$ are the values that estimate image intensities in Ω_i . K is a Gaussian kernel expressed as equation (15).

$$K_\sigma(u) = \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-|u|^2/2\sigma^2} \quad (15)$$

with scale parameter $\sigma > 0$.

To extract the object's whole boundary, \mathcal{E}_x^{Fit} is minimized by C for all x in the image domain. This is achieved by minimizing the integral of \mathcal{E}_x^{Fit} over all the center points as shown in equation (16).

$$\mathcal{E}(C, f_1(x), f_2(x)) = \int \mathcal{E}_x^{Fit}(C, f_1(x), f_2(x)) dx + \nu |C| \quad (16)$$

The zero-level set of the Lipschitz function $\phi : \Omega \rightarrow k$ which is known as a level set function represents $C \subset \Omega$. Let ϕ take the positive and negative values of C and H be the Heaviside. The energy function is expressed as equation (17).

$$\mathcal{E}_x^{Fit}(\phi, f_1(x), f_2(x)) = \sum_{i=1}^2 \lambda_i \int K_\sigma(x-y) |I(y) - f_i(x)|^2 M_i(\phi(y)) dy \quad (17)$$

where $M_1(\phi) = H(\phi)$ and $M_2(\phi) = 1 - H(\phi)$.

The energy \mathcal{E} in equation (16) can be redefined as equation (18).

$$\begin{aligned} \mathcal{E}_\epsilon(\phi, f_1, f_2) = & \sum_{i=1}^2 \lambda_i \int \left(\int K_\sigma(x-y) |I(y) - f_i(x)|^2 M_i(\phi(y)) dy \right) dx \\ & + \nu \int |\nabla H(\phi(x))| dx \end{aligned} \quad (18)$$

where $\int |\nabla H(\phi(x))| dx$ computes the zero-level contour of ϕ .

The Heaviside function H expressed in the equation above is estimated by a smooth function H_ϵ expressed as equation (19) and its derivative is a Dirac delta function equation (20).

$$H_\epsilon = \frac{1}{2} \left[1 + \frac{2}{\pi} \tan^{-1} \left(\frac{x}{\epsilon} \right) \right] \quad (19)$$

$$\delta_\epsilon(x) = H'_\epsilon(x) = \frac{1}{\pi} \frac{\epsilon}{\epsilon^2 + x^2} \quad (20)$$

By substituting H in equation (18) with H_ϵ , the definition of the energy function is then expressed as in equation (21).

$$\begin{aligned} \mathcal{E}_\epsilon(\phi, f_1, f_2) = & \sum_{i=1}^2 \lambda_i \int \left(\int K_\sigma(x-y) |I(y) - f_i(x)|^2 M_i^\epsilon(\phi(y)) dy \right) dx \\ & + \nu \int |\nabla H_\epsilon(\phi(x))| dx \end{aligned} \quad (21)$$

where $M_1^\epsilon(\phi) = H_\epsilon(\phi)$ and $M_2^\epsilon(\phi) = 1 - H_\epsilon(\phi)$.

A level set regularization term to preserve the regularity of level set function ϕ for accurate computation and stable level set evolution is introduced as equation (22) and therefore the energy function is expressed as equation (23).

$$P(\phi) = \int \frac{1}{2} (|\nabla \phi(x)| - 1)^2 dx \quad (22)$$

$$F(\phi, f_1, f_2) = \mathcal{E}_\epsilon(\phi, f_1, f_2) + \mu P(\phi) \quad (23)$$

For a fixed level set function ϕ we minimize equation (23) with

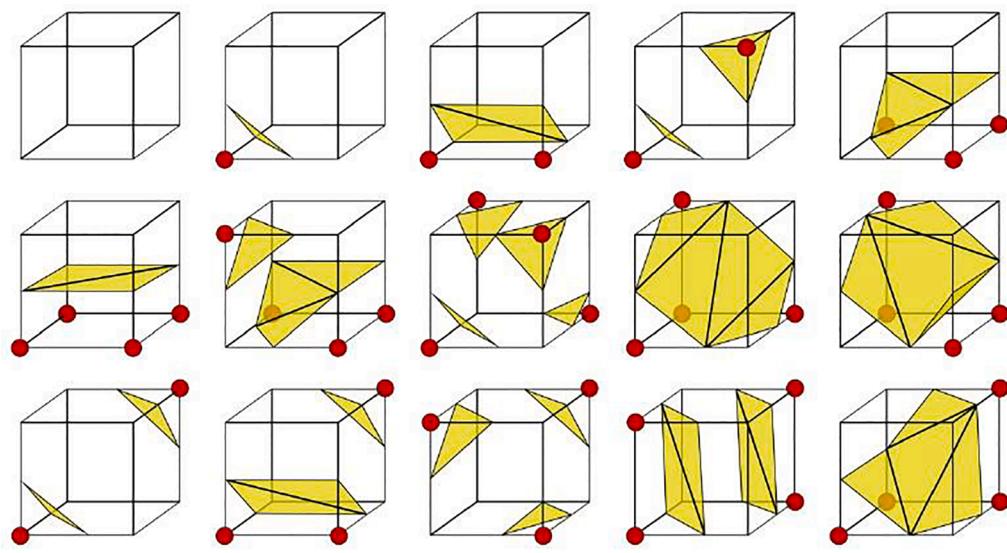


Fig. 5. Marching cubes 15 unique cases (Laprairie & Hamilton, 2021).

respect to $f_1(x)$ and $f_2(x)$ by calculus of variations to calculate equation (24).

$$\int K_\sigma(x-y) M_i^\epsilon(\phi(y)) (I(y) - f_i(x)) dy = 0, \quad i = 1, 2 \quad (24)$$

From equation (24) we obtain equation (25).

$$f_i(x) = \frac{K_\sigma(x)^* [M_i^\epsilon(\phi(x)) I(x)]}{K_\sigma(x)^* M_i^\epsilon(\phi(x))}, \quad i = 1, 2 \quad (25)$$

The function of $f_i(x)$ given by equation (25) are weighted averages of intensities in a neighborhood of x whose size is proportional to the scale parameter σ .

Keeping f_1 and f_2 static, the energy function $F(\phi, f_1, f_2)$ is minimized with respect to ϕ using the standard gradient descent method by working out the gradient flow equation as shown in equation (26).

$$\frac{\partial \phi}{\partial t} = -\delta_\epsilon(\phi)(\lambda_1 e_1 - \lambda_2 e_2) + \nu \delta_\epsilon(\phi) \operatorname{div}\left(\frac{\nabla \phi}{|\nabla \phi|}\right) + \mu \left(\nabla^2 \phi - \operatorname{div}\left(\frac{\nabla \phi}{|\nabla \phi|}\right) \right) \quad (26)$$

where δ_ϵ is the smoothed Dirac delta function in equation (20), e_1 and e_2 are expressed as equation (27).

$$e_i(x) = \int K_\sigma(x-y) |I(x) - f_i(y)|^2 dy, \quad i = 1, 2 \quad (27)$$

where f_i is given by equation (25).

Equation (26) is the objective function where $-\delta_\epsilon(\phi)(\lambda_1 e_1 - \lambda_2 e_2)$ is the fitting term, $\nu \delta_\epsilon(\phi) \operatorname{div}(\nabla \phi / |\nabla \phi|)$ is the arc length term and $\mu (\nabla^2 \phi - \operatorname{div}(\nabla \phi / |\nabla \phi|))$ is the level set regularization term.

2.2.4. Marching cubes

Marching cubes algorithm (Lorensen & Cline, 1987) was used to render an object into space. This is a 3D isosurface technique that creates triangular mesh from implicit functions by iterating (Marching) through linear array of cubes overlaid over an area. The isocontours are retrieved from the 3D entity's data in order to determine whether or not the isosurface will March through the unit cube. Then multi polygons are generated and isosurface is simulated. The procedure is described below.

1. Provide the isovalue.
2. Classify the unit cube by the isovalue.

3. Generate polygonal grids and reorganize the polygonal grids of each unit cube to complete the Marching cube algorithm.

Each cube has 8 vertices, and if all of them are positive or negative, the cube is entirely above or beneath the surface with no triangles emitted. Since each vertex can either be positive or negative, there are 256 possible configurations, but many are equivalent to each other. Fig. 5 show 15 unique cases. Iteration over all cubes is done by adding all triangles making a mesh through linear interpolation.

3. Experiment and results

The designed system experiment was conducted using the QIN lung CT dataset. We used case R 0273 for demonstration; slices: 12, slice thickness: 5 mm, pixel spacing: 0.8046875.

3.1. Evaluation metrics

To measure the capability of the detection system, we employed Intersection over Union (IoU) which evaluates the intersection of the predicted bounding box with the ground truth's bounding box of the detected object (Rezatofighi et al., 2019). If the IoU value is big, it implies that the overlaps between the predicted bounding box and the ground truth bounding box is huge. The mathematical expression of an IoU is shown in equation (28).

$$IoU = \frac{TP}{FN + TP + FP} = \frac{A \cap B}{A \cup B} \quad (28)$$

where TP , FN , FP is true positive, false negative and false positive respectively.

Precision and sensitivity were used to evaluate the localization of the proposed system. We used precision to determined how many of the anticipated positives were identified correctly and with sensitivity we determined how many of the true positives were properly identified. Both precision and sensitivity are calculated as show below in equation (29) and (30) respectively.

$$precision = \frac{TP}{TP + FP} \quad (29)$$

$$sensitivity = \frac{TP}{TP + FN} \quad (30)$$

We also utilized the F 1 score to calculate the harmonic mean

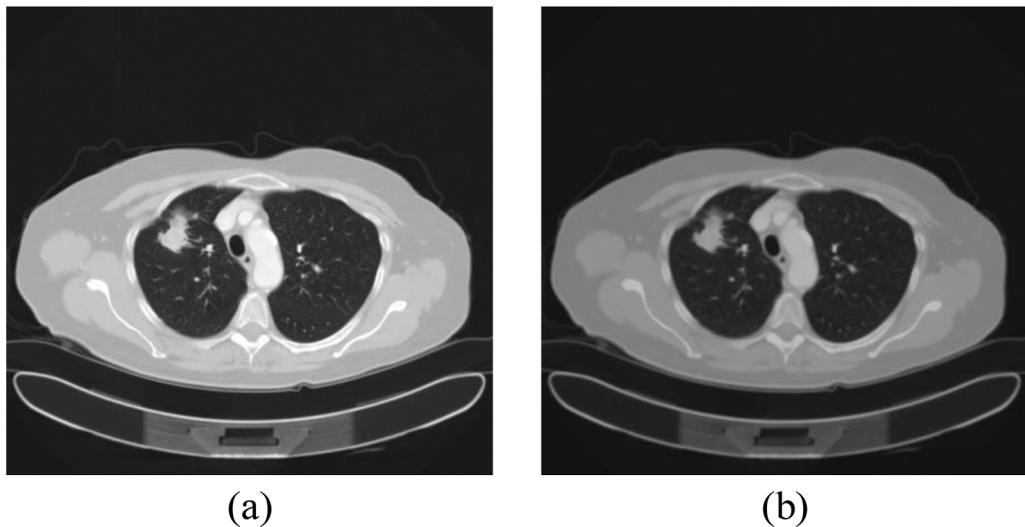


Fig. 6. Noise reduction filtering; (a) original image, (b) Mean filter.

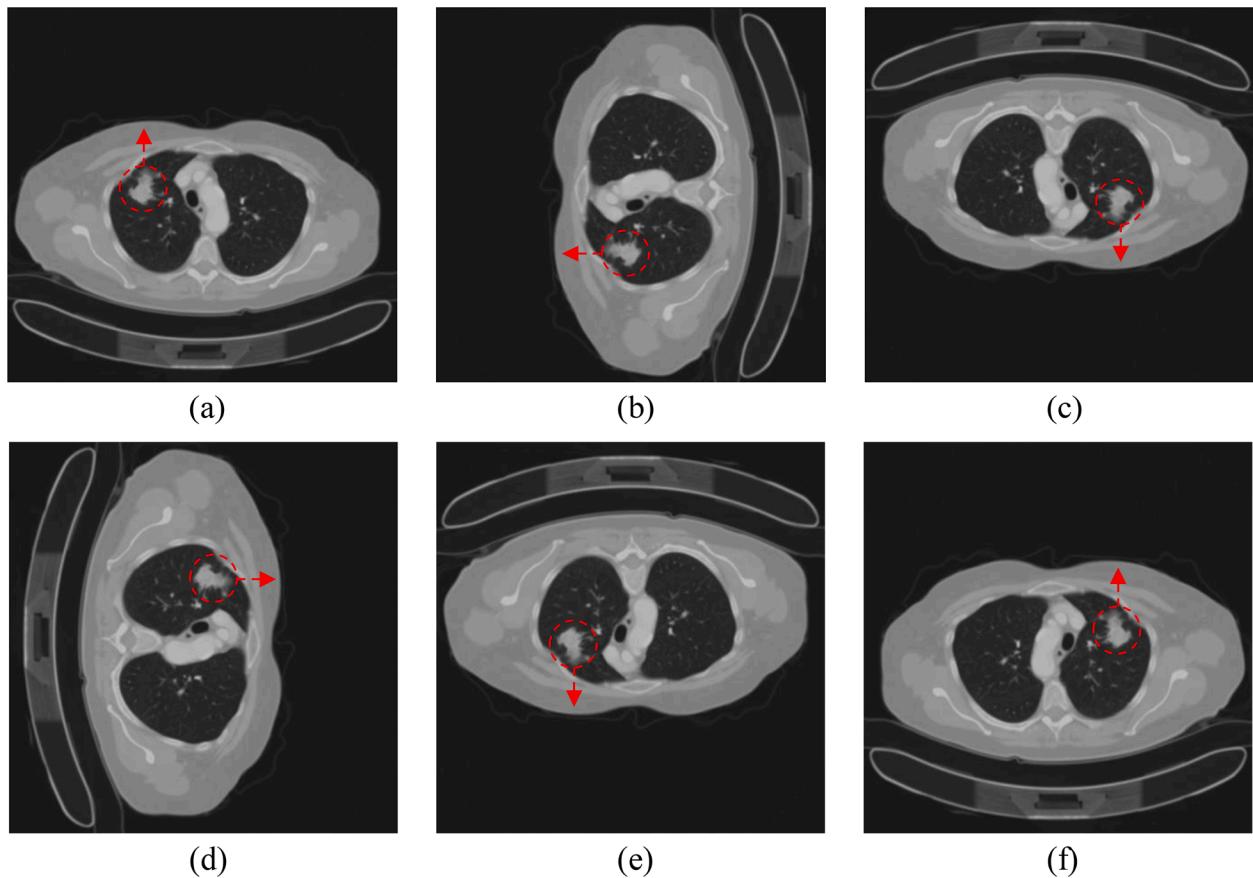


Fig. 7. Affine transformation augmentation; (a) original, (b) 90° rotation, (c) 180° rotation, (d) 270° rotation, (e) horizontal flip, (f) vertical flip.

between precision and sensitivity. Since the F_1 score is a mean of precision and sensitivity, it means that it always gives equal weight to both measures. This is because there's always a trade-off between the two, as the mean value is always greatly affected at the expense of either value. F_1 score equation is shown in equation (31) below.

$$F1\text{ score} = \frac{2 \cdot \text{precision} \cdot \text{sensitivity}}{\text{precision} + \text{sensitivity}} \quad (31)$$

A further evaluation was done with mean Average Precision (mAP)

to evaluate the localization accuracy. mAP is an object detection measure by Microsoft Common Object in Context (COCO) (Lin et al., 2014) which works by averaging over multiple IoU thresholds averaged over all categories. $AP[0.5:0.05:0.95]$ corresponds to the average AP for IoU from 0.5 to 0.95 with step size of 0.05. Considering the fact that we only focused on just finding the tumor, we used $mAP@.5$ as an evaluation criterion in this study.

The tumor extraction capability of the system was measured using a dice score coefficient (DSC) which measures the intersection between

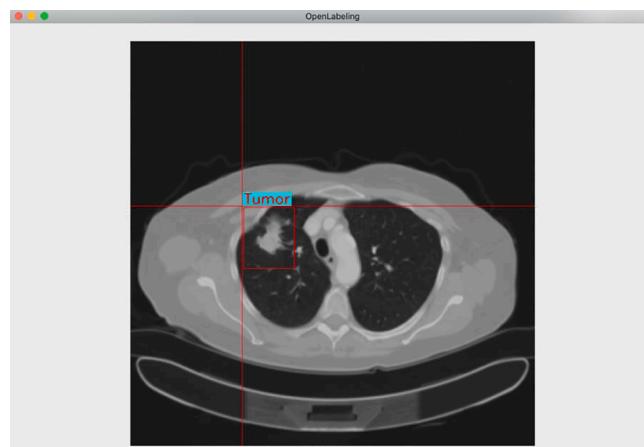


Fig. 8. Manual labeling.

the system generated segments and the ground truth, divided by the summation of the segments and ground truth. Equation (32) shows the scientific expression of the DSC.

$$DSC = \frac{2TP}{2TP + FN + FP} = \frac{2|A \cap B|}{|A| + |B|} \quad (32)$$

3.2. Image enhancement

The quality of the data was poor since the images were corrupted with CT noise. To eliminate the noise, a one-time pass 3×3 kernel Mean filter was utilized using a stride of 1. A sample of the enhanced images is shown in Fig. 6 (b), compared side to side with the original image.

3.3. Augmentation

After the images were filtered, we augmented them to diversify and multiply the data. The images were transformed into five affine transformations with light intensity of each transformation set to high and low. This implies that we ended up with more than 18 different sorts of images from a single source of image. This strategy encourages the

model to deal with a wider range of tumor at different light intensities. Fig. 7 shows the affine transformation of the augmentation with reference to the tumor (arrow pointing torso direction).

3.4. Labeling

After the augmentation, we used OpenLabeling, a freely available python based graphical labeling source code for labeling images to label the tumor nodules in our dataset. With the assistance of professional clinicians, we manually labeled the tumor nodules with bounding boxes and tagged them with a label as shown in Fig. 8. The process was, one clinician will label and tag each subject and pass it over to another for verification. The labeled images were then saved as *.txt files to be utilized in the localization training later.

3.5. Localization

The localization experiment was conducted using Python 3.7 platform on a Windows 10 workstation Intel (R) Core i 5 –4430, 147 GB RAM and an Nvidia Tesla P 100 GPU using Cuda 11.0 toolkit.

The model was trained in two sessions to maximize its localization potential. First, we pre-trained the model from scratch with 3000 iterations using a batch size of 32, learning rate of 0.0001, momentum of 0.949 and decay of 0.005. We achieved a training loss of 3.751. The data was divided into three partitions with the ratio of 7 : 2 : 1 for train, valid and test respectively. In the second training session, we used the best weights from the pre-trained model to train for 2000 more iterations using the same hyper parameters resulting in a training loss of 0.0213 with a validation loss of 9.681 as shown in Fig. 9 below. Table 1 displays

Table 1
Localization training hyper parameters.

| Hyper parameters | Tuning |
|-------------------------|--------|
| Pre-training iterations | 3000 |
| Training iterations | 2000 |
| Learning rate | 0.0001 |
| Batch size | 32 |
| Momentum | 0.949 |
| Decay | 0.005 |

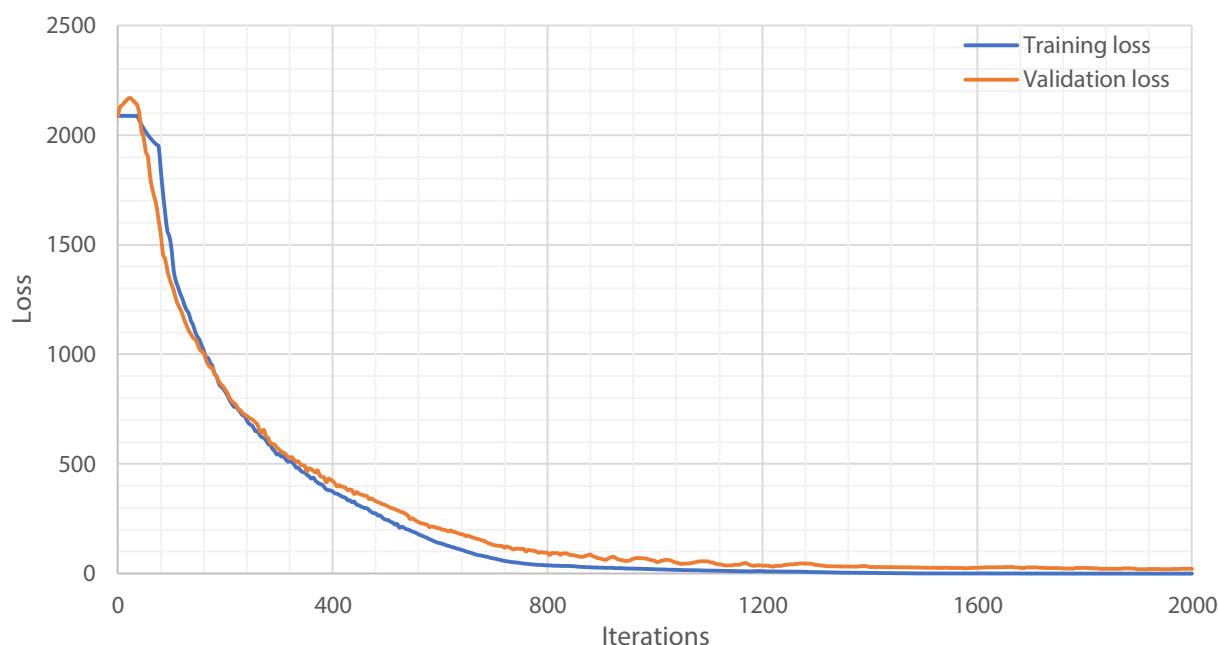


Fig. 9. Training and validation loss.

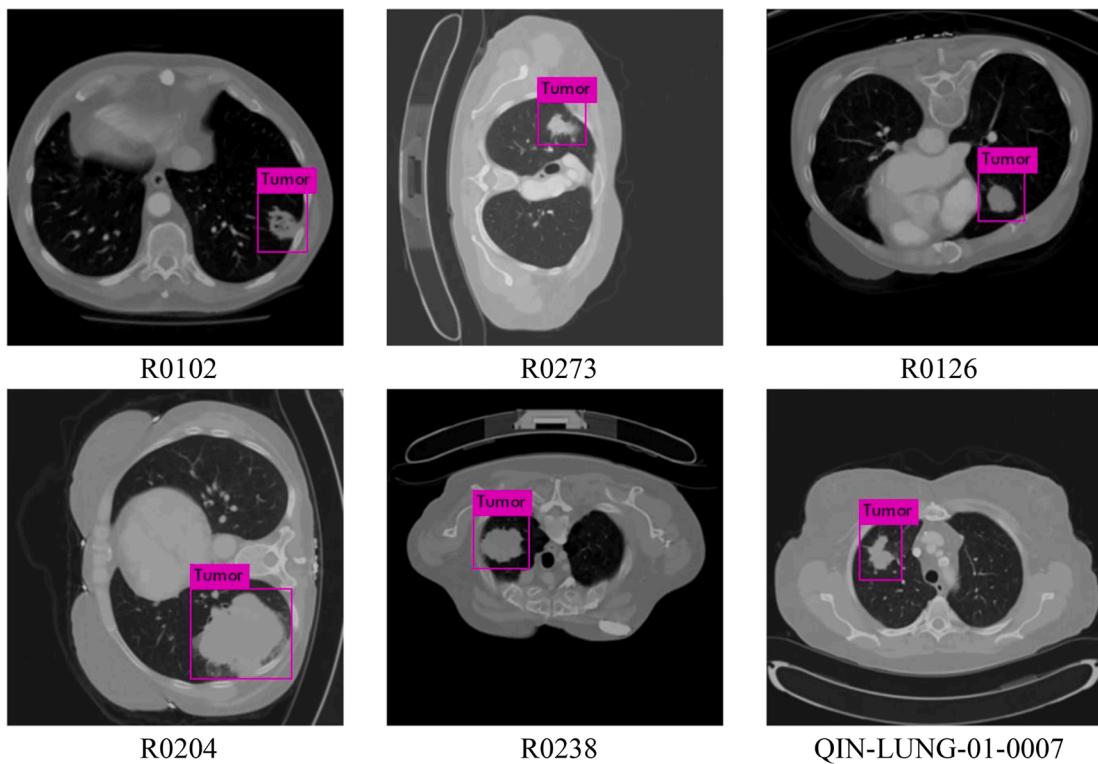


Fig. 10. QIN dataset test results.

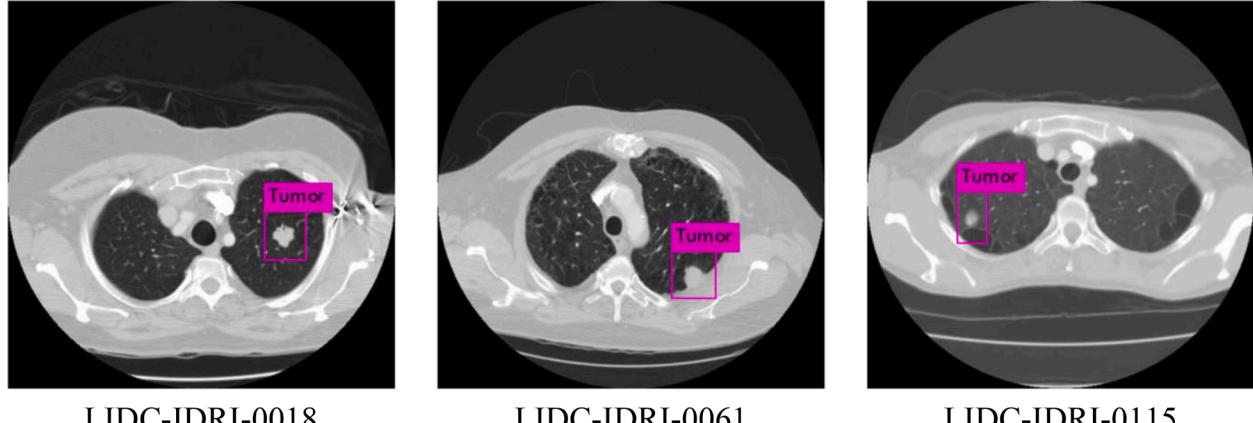


Fig. 11. LIDC-IDRI dataset test results.

all the hyper parameter tunings for our localization model.

The tumor detection achieved an $mAP@.5$ of 88.76% at 34 fps in our GPU. Detection precision and sensitivity was 96.57% and 97.02% respectively with an $F1$ score of 96.79%. The achieved average IoU was 92.34% at 21.38 ms average detection time per image. A random pick of the results from their respective cases in QIN dataset is shown in Fig. 10. Additional test results from LIDC-IDRI dataset is shown in Fig. 11 demonstrating the efficacy and reliability of our detection model.

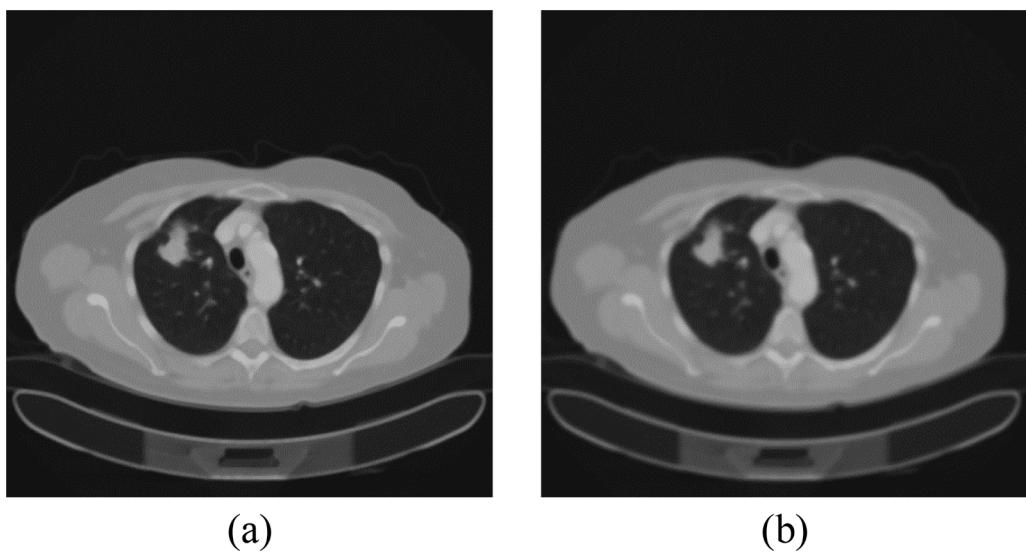
3.6. Image filtering

In order to build a 3D model of the detected tumor nodules, a Median filter was employed to eliminate impulsive noise while preserving edges and lines. The Median filter was used with a one-time pass, 5×5 kernel using stride 1. To blur the images, a Gaussian 3×3 filter was single passed with a stride of 1. The filtered images are presented in Fig. 12.

According to the results, noise was removed with edges preserved and the images were successfully blurred for further processing.

3.7. Tumor extraction

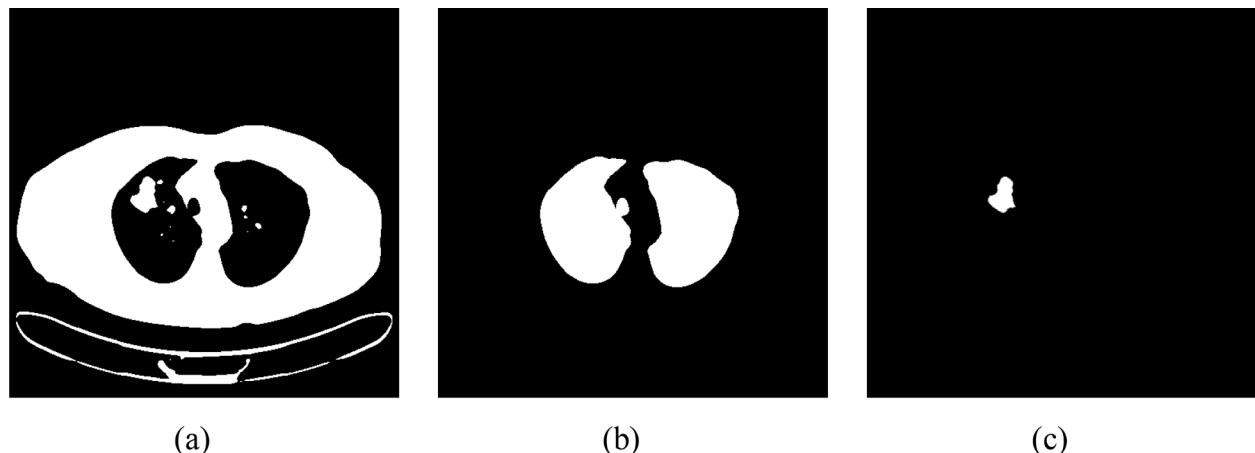
When image filtering was completed, the tumor was extracted by using k -means and morphology. k -means was employed to obtain a binary cluster of the image by pixel clustering. According to Krishna-murthy et al. (2017) the parameter setting used in the clustering were; a cluster center of 2, maximum iterations of 10 and epsilon of 1 for accuracy. The gray levels were divided into 2 clusters in the image and they were labeled respectively. We then employed Otsu threshold to binarize the image as shown in Fig. 13(a). The lung mask was obtained by morphology of erosion, opening and dilation and connecting labels. Results shown in Fig. 13(b). To extract the tumor mask, the clustered image was multiplied with the lung mask, and the remaining vascular



(a)

(b)

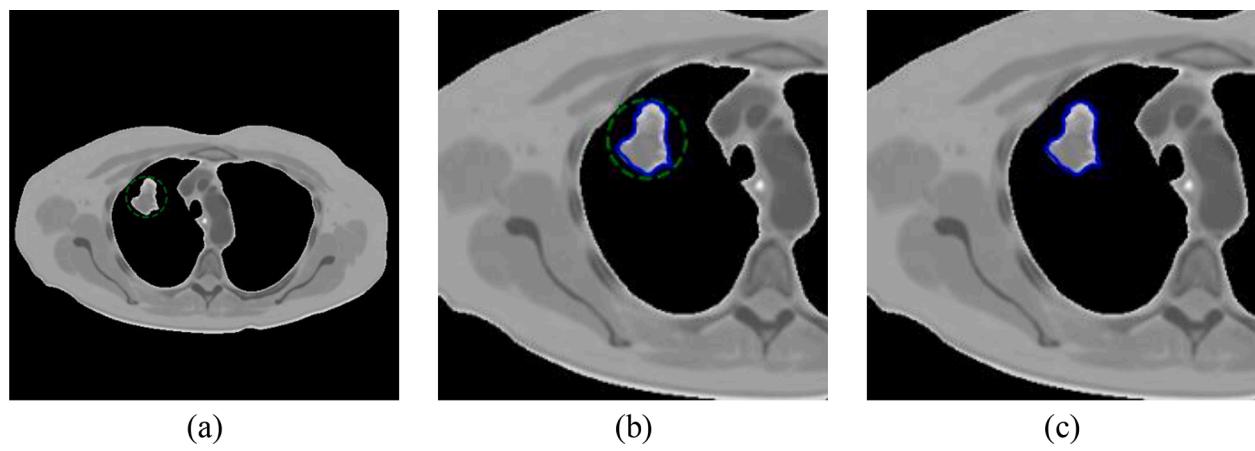
Fig. 12. Image filtering; (a) Median filter, (b) Gaussian blur.



(a)

(b)

(c)

Fig. 13. Tumor extraction; (a) k -means clustering, (b) lung mask, (c) tumor mask.

(a)

(b)

(c)

Fig. 14. Region-based active contour; (a) region descriptor, (b) guided active contour's motion, (c) extracted contour.

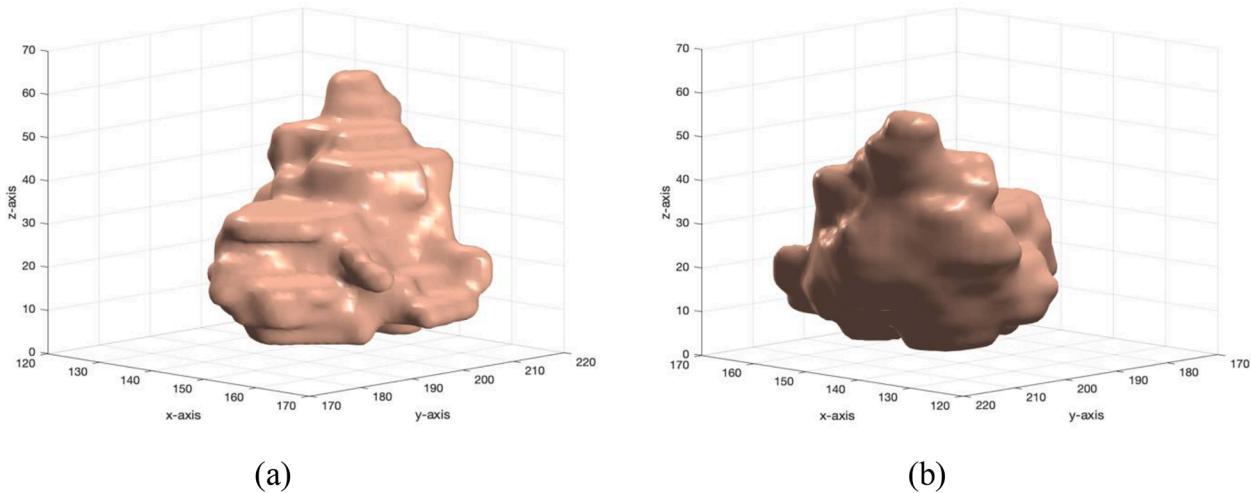


Fig. 15. 3D Tumor reconstruction result; (a) front view, (b) back view.

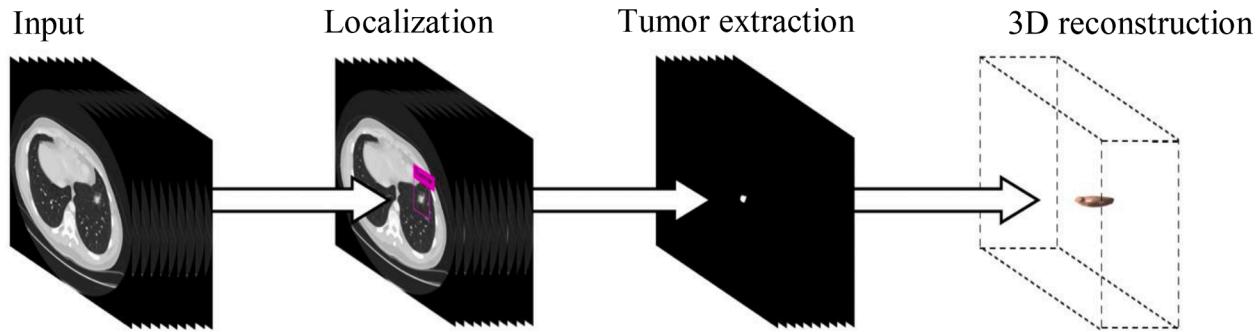


Fig. 16. System schematics: input, localization, tumor extraction, 3D reconstruction.

points were eliminated by morphology and connecting labels. The extracted tumor mask is shown in Fig. 13(c).

3.8. Region-based active contour

To extract the contour information of the tumor nodules we used region-based active contour. This model requires a region descriptor of the region of interest to identify the tumor nodule boundary.

The knowledge of the position for the localized tumor is key for accurate estimation of the nodule boundary. According to Albahli et al. (2020), we used the bounding box position and specified the region descriptor of the region of interest inside its proximity. This region descriptor guides the active contour's motion to extract the contours of the tumor by approximating the intensity of the image inside the defined region which makes it possible to perfectly capture the contour of a rough tumor shape. Fig. 14 shows the region-based active contour results. To initiate automation, the contour center of the previous tumor is utilized as a seed of the next image's region of interest contributing to the growth of the next tumor in the next image, and so forth.

3.9. 3D reconstruction

Marching cubes algorithm was used in the volumetric rendering of the tumor into a 3D. The contour information extracted by the region-based active contours were stacked in a 3D matrix data which was projected into space. Each element in this space was defined as a voxel and the isosurface of each was calculated. The 3D volumetric rendering result is shown in Fig. 15 where the x-axis and y-axis represents the 3D x-pixels and y-pixels of the image respectively and the z-axis is the pixel height of the tumor.

Table 2
Tumor extraction validation.

| Sample No. | Samples | No of slices | DSC |
|------------|----------------|--------------|--------|
| 1 | LIDC-IDRI-0001 | 9 | 0.9082 |
| 2 | LIDC-IDRI-0018 | 6 | 0.9248 |
| 3 | LIDC-IDRI-0023 | 11 | 0.9129 |
| 4 | LIDC-IDRI-0043 | 7 | 0.9204 |
| 5 | LIDC-IDRI-0050 | 10 | 0.9347 |
| 6 | LIDC-IDRI-0057 | 21 | 0.9113 |
| 7 | LIDC-IDRI-0061 | 14 | 0.9279 |
| 8 | LIDC-IDRI-0089 | 10 | 0.9292 |
| 9 | LIDC-IDRI-0115 | 8 | 0.9145 |
| 10 | LIDC-IDRI-0127 | 13 | 0.9351 |

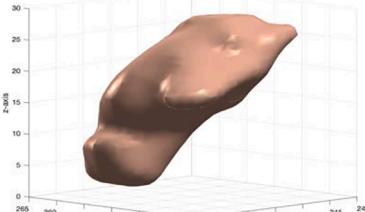
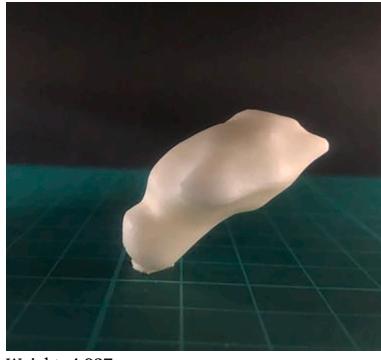
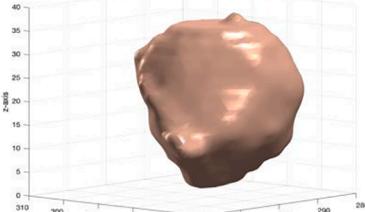
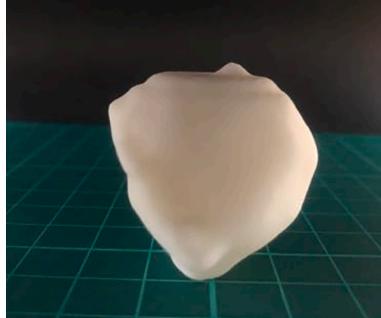
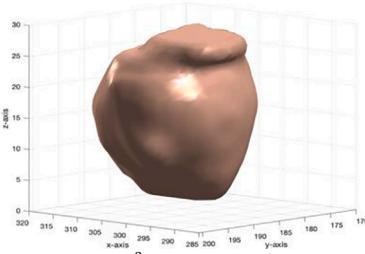
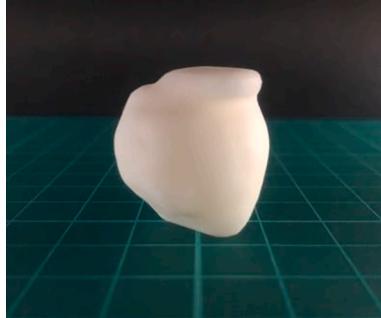
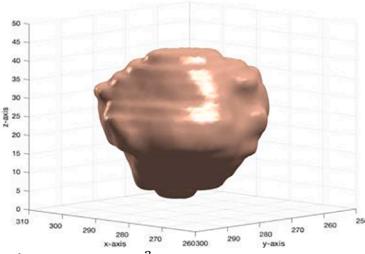
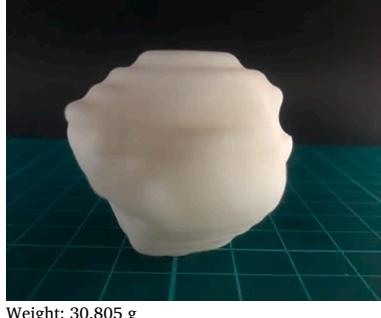
4. System validation

The validation of the proposed system was conducted to test the system capability to extract accurate tumor masks and render an accurate volumetric tumor using a different unseen data. Simplified schematics of the system is shown in Fig. 16.

In order to validate the system tumor extraction capability, we tested the system with LIDC-IDRI dataset. We used DSC to evaluate the segmentation accuracy whereby we measured the segmentation overlaps between the extracted tumor masks and the ground truth divided by the summation of both elements. The mathematical expression of DSC is shown in equation (32) where the numerator is the intersection multiplied by 2 over the summation of the union and intersection. We randomly selected 10 test samples from the LIDC-IDRI data. The ground truth masks of the test samples were manually prepared by professional clinicians and verified by a thoracic surgeon (Y-H. Chen with more than

Table 3

Volumetric tumor validation.

| | 3D rendered image | 3D printed image |
|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| LIDC-IDRI-0057Pixel spacing: 0.730469 Slice thickness: 1.25 No. of slices: 21 |  <p>Volume: 4.016 cm³</p> |  <p>Weight: 4.827 g Volume: 4.023 cm³</p> |
| LIDC-IDRI-0061Pixel spacing: 0.820312 Slice thickness: 2.5 No. of slices: 14 |  <p>Volume: 17.317 cm³</p> |  <p>Weight: 20.714 g Volume: 17.262 cm³</p> |
| LIDC-IDRI-0089Pixel spacing: 0.820312 Slice thickness: 2.5 No. of slices: 10 |  <p>Volume: 8.051 cm³</p> |  <p>Weight: 9.681 g Volume: 8.068 cm³</p> |
| LIDC-IDRI-0127Pixel spacing: 0.779297 Slice thickness: 3 No. of slices: 13 |  <p>Volume: 25.777 cm³</p> |  <p>Weight: 30.805 g Volume: 25.671 cm³</p> |

10 years of experience in thoracic CT interpretation) from Tri-Service General Hospital in Taipei, Taiwan. The system achieved an average DSC of 92.19%. Table 2 shows the DSC per case of the 10 randomly selected test cases.

The volumetric tumor was validated by printing out the rendered 3D tumor images to verify their volume. We chose 4 samples at random to

verify the accuracy of the volume estimation. The actual volumes acquired with our method during testing are presented in Table 3. They were acquired by using the active contour method presented in the study where each tumor nodule was calculated as a cross-section and multiplied by its thickness. The 3D images were printed exactly the same as the actual tumors obtained in the study method. The weights of the

Table 4
Comparison with YOLO predecessors.

| Study | Model | Dataset | Sensitivity |
|----------------------|--------|-------------|-------------|
| George et al. (2018) | YOLO | LIDC-IDRI | 89.0 % |
| Aresta et al. (2018) | YOLOv2 | KDSB17 | 89.4 % |
| Xu et al. (2020) | YOLOv3 | LUNA16 | 92.6 % |
| Ours | YOLOv4 | QIN LUNG CT | 96.8 % |

Table 5
Comparison with other related detection studies.

| Study | Model | Data | Sensitivity |
|--------------------------|------------------------|---------------|-------------|
| Zhang et al. (2019) | Deep CNN | LUNA16 & KDSB | 84.4 % |
| Sori et al. (2019) | Multi-path CNN | KDBS 17 | 87.4 % |
| Kasinathan et al. (2019) | Enhanced CNN | LIDC-IDRI | 89 % |
| Sang et al. (2019) | 3D-MixNet-Faster R-CNN | LIDC-IDRI | 94 % |
| Ours | YOLOv4 | QIN LUNG CT | 96.8 % |

printed tumors are presented in [Table 3](#) as well. The printing material used was a standard resin with a density of 1.20 g.cm^{-3} . We used the density formula ($\text{density} = \text{mass/volume}$) to calculate the volumes of the printed images. Our method displayed a high reliability obtaining an overall error difference of 0.137 cm^3 with a percentage error of 0.258 %.

5. Discussion

Lung tumor volumetric analysis is essential in medical practice to give doctors a clear information of the tumor to assist them in executing a proper planning of patient's treatment. Unlike other industries where human error can be considered as collateral and easily absorbed by the company, in this field, it can put a life in danger, worse even death. High-level of details is therefore a necessary component to avoid unnecessary complications.

In this study, a system that detects, segments and accurately renders NSCLC tumor volumetric automatically using YOLOv4 and region-based active contour model was proposed. The design of this system is made out of two sections which are the detection and the volumetric rendering. The detection is composed of image enhancement, augmentation, labeling and localization while the volumetric rendering is mainly composed of image filtering, tumor extraction, region-based active contour and 3D reconstruction. This study presented all details and results on all implementations conducted at every level of the proposed system.

CT images are often distorted by a great deal of noise, therefore the first process in this study was to remove the noise and improve the images quality using a Mean filter. We then used the technique of data augmentation to diversify the image data to create a flexible and robust detection model. Before the training of the localization model, a python based labeling tool was utilized to tag and label all the tumors in the dataset. After a successful training and testing, the detected images with localized tumors were filtered using a Median filter and Gaussian filter to smooth out any noise while preserving important information on the images. k -means clustering was brought in to cluster pixels with similar properties before the use of binary threshold and morphology to extract tumor masks. The volumetry was rendered with marching cubes after extracting the contour information using region-based active contour.

Medical image analysis can be very complicated especially due to the complex structure of the images used in medicine. A good image pre-processing with a good understanding of the images can lead to a good overall model. The technique we applied in the use of YOLOv4 in our study mostly countered the accuracy and the speed issue experienced by a lot of detection models. We trained a deep YOLOv4 model from our pretrained weights we trained from scratch and achieved a detection

Table 6
Comparison with other related segmentation studies.

| Study | Model | Data | DSC |
|----------------------|-------------------------|-----------|---------|
| Kamal et al. (2020) | Recurrent 3D Dense Unet | LIDC-IDRI | 72.28 % |
| Huang et al. (2019) | Deep CNN | LIDC-IDRI | 79.3 % |
| Koç & Güveniş (2020) | AC & Random Forest | LIDC-IDRI | 87.3 % |
| Ours | YOLOv4 & RBAC | LIDC-IDRI | 92.19 % |

AC: active contour, RBAC: region-based active contour.

precision, sensitivity and F 1 score of 96.57%, 97.02% and 96.79% respectively at 34 fps in our GPU. We achieved a real-time detection speed of 21.38 ms per image. To evaluate the performance of our model, it was compared to its YOLO predecessors in lung cancer detection as shown in [Table 4](#). It is of no doubt that it out performed its antecedent in all aspects due to the architectural upgrades and the training techniques we used. Although [Xu et al. \(2020\)](#) put out a competitive sensitivity performance with a detection speed of 31.3 ms, but still came short on precision at 83.5%. A further comparison with other detection models in related studies on the same field was conducted as well. Our model showed a dominance of 2.8% over [Sang et al. \(2019\)](#) 3D-MixNet-Faster R-CNN detection model while at the same time showing a more than 10% performance over [Zhang et al. \(2019\)](#) Deep CNN as shown in [Table 5](#).

The system's rendering performance was also evaluated using LIDC-IDRI as test dataset. It was evaluated by its capability to segment and extract tumors similar to its ground truth. The system displayed a good performance of 92.19% DSC. The tumor extraction capability was compared to other related studies as shown in [Table 6](#).

The volumetric rendering was verified by printing out 3D images of the rendered tumors in the study and compare their volumes side by side. We achieved near perfect results of 99.74 % accuracy. The system took about 11 s average time to render a complete volumetric tumor. The speed performance depends on a lot of things, the obvious one being the data series size. This study showed competitiveness when compared to other studies used the same verification method. [Dang et al. \(2013\)](#) achieved 96% accuracy at 1 min 20 s on a realistic brain phantom using the same approach. Better results were obtained by [Kuo et al. \(2017\)](#) achieving an accuracy of 98.08% at 7.25 seconds on advanced lung cancer detection. Our designed system definitely out performs its competitors, although falling short in rendering time, it still shows dominance and strong reliability in the volume accuracy.

6. Conclusion

The aim of the system presented in this study is to automatically detect, segment and reconstruct an accurate 3D tumor volumetric for NSCLC to assist doctors in the planning of surgical procedures and assessing patient's treatment response. Localization is obtained by the cutting edge YOLOv4 convolutional neural network which achieved outstanding results in real-time detection. k -means and region-based active contour played a vital role in obtaining contour information to render the volumetric tumor into space utilizing marching cubes algorithm. We built and tested our system using publicly available TCIA dataset. The rendered tumor was verified by printing a 3D image of the actual tumor which achieved near perfect match. The process from localization to rendering can take an average time of about 11 s depending on the series size.

In the future, a clinical verification needs to be done to confirm the contribution of the presented study in clinical practice. Furthermore, if possible, for surgical patients, the designed model's results must be compared with the post-operative actual tumor volume and pathological tests must be done to provide a criterion for malignant and benign images.

CRediT authorship contribution statement

Sifundvolesihle Dlamini: Investigation, Writing – original draft. **Yi-Hsi Chen:** Investigation, Methodology. **Chung-Feng Jeffrey Kuo:** Conceptualization, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The research was supported by the Tri-Service General Hospital, National Defense Medical Center-National Taiwan University of Science and Technology Joint Research Program (TSGH-A-111-004, TSGH-NTUST-111-03).

References

- Albahli, S., Nida, N., Irtaza, A., Yousaf, M. H., & Mahmood, M. T. (2020). Melanoma lesion detection and segmentation using YOLOv4-DarkNet and active contour. *IEEE Access*, 8, 198403–199414. <https://doi.org/10.1109/ACCESS.2020.3035345>
- Aresta, G., Araújo, T., Jacobs, C., van Ginneken, B., Cunha, A., Ramos, I., & Campilho, A. (2018). Towards an automatic lung cancer screening system in low dose computed tomography. In *Image Analysis for Moving Organ, Breast, and Thoracic Images*, Springer, (pp. 310–318).
- Armato, S. G., III, McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., ... Hoffman, E. A. (2011). The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans. *Medical Physics*, 38(2), 915–931. <https://doi.org/10.1118/1.3528204>
- Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y.M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.
- Bouveyron, C., Girard, S., & Schmid, C. (2007). High-dimensional data clustering. *Computational Statistics & Data Analysis*, 52(1), 502–519. <https://doi.org/10.1016/j.csda.2007.02.009>
- Caو, Z., Liao, T., Song, W., Chen, Z., & Li, C. (2021). Detecting the shuttlecock for a badminton robot: A YOLO based approach. *Expert Systems with Applications*, 164, Article 113833. <https://doi.org/10.1016/j.eswa.2020.113833>
- Cartucho, J., Ventura, R., & Veloso, M. (2018). Robust object recognition through symbiotic deep learning in mobile robots. In *In 2018 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, IEEE (pp. 2336–2341). <https://doi.org/10.1109/IROS.2018.8594067>
- Chan, T. F., & Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on Image Processing*, 10(2), 266–277. <https://doi.org/10.1109/83.902291>
- Chen, Y., Han, C., Wang, N., & Zhang, Z. (2019). Revisiting feature alignment for one-stage object detection. *arXiv preprint arXiv:1908.01570*.
- Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., ... Pringle, M. (2013). The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging*, 26(6), 1045–1057. <https://doi.org/10.1007/s10278-013-9622-7>
- Dang, M., Modi, J., Roberts, M., Chan, C., & Mitchell, J. R. (2013). Validation study of a fast, accurate, and precise brain tumor volume measurement. *Computer Methods and Programs in Biomedicine*, 111(2), 480–487. <https://doi.org/10.1016/j.cmpb.2013.04.011>
- Deng, G., & Cahill, L. (1993). An adaptive Gaussian filter for noise reduction and edge detection. In *In 1993 IEEE conference record nuclear science symposium and medical imaging conference. IEEE* (pp. 1615–1619). <https://doi.org/10.1109/NSSMIC.1993.373563>
- Dhanachandra, N., Manglem, K., & Chanu, Y. J. (2015). Image segmentation using K-means clustering algorithm and subtractive clustering algorithm. *Procedia Computer Science*, 54, 764–771. <https://doi.org/10.1016/j.procs.2015.06.090>
- Krishnamurthy, S., Narasimhan, G., & Rengasamy, U. Lung nodule growth measurement and prediction using auto cluster seed k-means morphological segmentation and shape variance analysis. *J. Biomed. Eng.*, 24(1), 53–71. 10.1504/IJBET.2017.083818.
- Dogra, A., & Bhalla, P. (2014). Image sharpening by gaussian and butterworth high pass filter. *Biomedical and Pharmacology Journal*, 7(2), 707–713. <https://doi.org/10.13005/bpj/545>
- George, J., Skaria, S., & Varun, V. (2018). Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans. In *In Medical Imaging 2018: Computer-Aided Diagnosis* (pp. 347–355). <https://doi.org/10.1117/12.2293699>
- Ghiasi, G., Lin, T.-Y., & Le, Q.V. (2018). Dropblock: A regularization method for convolutional networks. *arXiv preprint arXiv:1810.12890*.
- Godoy, M. C., Kim, T. J., White, C. S., Bogoni, L., De Groot, P., Florin, C., ... Naidich, D. P. (2013). Benefit of computer-aided detection analysis for the detection of subsolid and solid lung nodules on thin-and thick-section CT. *American Journal of Roentgenology*, 200(1), 74–83. <https://doi.org/10.2214/AJR.11.7532>
- Ghosh, S., & Dubey, S. K. (2013). Comparative analysis of k-means and fuzzy c-means algorithms. *International Journal of Advanced Computer Science and Applications*, 4(4), 35–39.
- Gupta, G. (2011). Algorithm for image processing using improved median filter and comparison of mean, median and improved median filter. *International Journal of Soft Computing and Engineering*, 1(5), 304–311.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9), 1904–1916. <https://doi.org/10.1109/TPAMI.2015.2389824>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).
- Hemalatha, R., Thamizhvanii, T., Dhivya, A. J. A., Joseph, J. E., Babu, B., & Chandrasekaran, R. (2018). Active contour based segmentation techniques for medical image analysis. *Medical and Biological Image Analysis*, 4(17), 2. <https://doi.org/10.5772/interchopen.74576>
- Huang, X., Sun, W., Tseng, T.-L.-B., Li, C., & Qian, W. (2019). Fast and fully-automated detection and segmentation of pulmonary nodules in thoracic CT scans using deep convolutional neural networks. *Computerized Medical Imaging and Graphics*, 74, 25–36. <https://doi.org/10.1016/j.compmedimag.2019.02.003>
- Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *In International conference on machine learning* (pp. 448–456).
- Jacobson, F.L., & Jaklitsch, M.T. (2018). Computed tomography scanning for early detection of lung cancer. *Annu. Rev.*, 69, 235–245. <https://doi.org/10.1146/annurev-med-020917-053556>
- Kamal, U., Rafi, A. M., Hoque, R., Wu, J., & Hasan, M. K. (2020). Lung cancer tumor region segmentation using recurrent 3d-denseunet. In *In International Workshop on Thoracic Image Analysis* (pp. 36–47).
- Kasinathan, G., Jayakumar, S., Gandomi, A. H., Ramachandran, M., Fong, S. J., & Patan, R. (2019). Automated 3-D lung tumor detection and classification by an active contour model and CNN classifier. *Expert Systems with Applications*, 134, 112–119. <https://doi.org/10.1016/j.eswa.2019.10.005>
- Katkar, J., Baraskar, T., & Mankar, V. R. (2015). A novel approach for medical image segmentation using PCA and K-means clustering. In *In 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, IEEE (pp. 430–435). <https://doi.org/10.1109/iCAT.2015.7456922>
- Koç, A., & Güveniç, A. (2020). Design and evaluation of an accurate CNR-guided small region iterative restoration-based tumor segmentation scheme for PET using both simulated and real heterogeneous tumors. *Medical and Biological Engineering and Computing*, 58(2), 335–355.
- Kuo, C.-F.J., Ke, B.-H., Wu, N.-Y., Kuo, J., & Hsu, H.-H. (2017). Prognostic value of tumor volume for patients with advanced lung cancer treated with chemotherapy. *Comput. Methods Programs Biomed.*, 144, 165–177. 101016/j.cmpb.2017.03.021.
- Lakshmanaprabu, S., Mohanty, S. N., Shankar, K., Arunkumar, N., & Ramirez, G. (2019). Optimal deep learning model for classification of lung cancer on CT images. *Future Gener. Comput. Systems*, 92, 374–382. <https://doi.org/10.1016/j.future.2018.10.009>
- Laprairie, M. J., & Hamilton, H. J. (2021). Isovox: A Brick-Octree Approach to Indirect Visualization. *Department of Computer Science*. ReasearchGate: University of Regina.
- Li, C., Kao, C.-Y., Gore, J. C., & Ding, Z. (2008). Minimization of region-scalable fitting energy for image segmentation. *IEEE Transactions on Image Processing*, 17(10), 1940–1949.
- Li, C., Xu, C., Gui, C., & Fox, M. D. (2010). Distance regularized level set evolution and its application to image segmentation. *IEEE Transactions on Image Processing*, 19(12), 3243–3254. <https://doi.org/10.1109/TIP.2010.2069690>
- Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117–2125).
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *In European conference on computer vision* (pp. 740–755).
- Li, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path aggregation network for instance segmentation. In *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8759–8768).
- Lorensen, W.E., & Cline, H.E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. *ACM Siggraph Computer Graphics* 21(4), 163–169. 101145/37402.37422.
- Ma, D., Liao, Q., Chen, Z., Liao, R., & Ma, H. (2019). Adaptive local-fitting-based active contour model for medical image segmentation. *Signal Processing: Image Communication*, 76, 201–213. <https://doi.org/10.1016/j.image.2019.05.006>
- Marten, K., Auer, F., Schmidt, S., Kohl, G., Rummery, E. J., & Engelke, C. (2006). Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria. *European Radiology*, 16(4), 781–790. <https://doi.org/10.1007/s00330-005-0036-x>
- Mathew, C.J., David, A.M., & Mathew, C.M.J. (2020). Artificial Intelligence and its future potential in lung cancer screening. *Excli. J.*, 19, 1552. 10.17179/excli2020-3095.
- Mozley, P. D., Schwartz, L. H., Bendtsen, C., Zhao, B., Petrick, N., & Buckler, A. J. (2010). Change in lung tumor volume as a biomarker of treatment response: A critical review

- of the evidence. *Annals of Oncology*, 21(9), 1751–1755. <https://doi.org/10.1093/annoc/mdq051>
- Mumford, D. B., & Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*. <https://doi.org/10.1002/cpa.3160420503>
- Nasser, I. M., & Abu-Naser, S. S. (2019). Lung cancer detection using artificial neural network. *International Journal of Engineering*, 3(3), 17–23.
- Nature portfolio. (2021). Lung cancer. <https://www.nature.com/subjects/lung-cancer> (Accessed: August 2021).
- Neal Joshua, E. S., Bhattacharyya, D., Chakravarthy, M., & Byun, Y.-C. (2021). 3D CNN with visual insights for early detection of lung cancer using gradient-weighted class activation. *Journal of Healthcare Engineering*. <https://doi.org/10.1155/2021/6695518>
- Neubeck, A., & Van Gool, L. (2006). Efficient non-maximum suppression. In: 18th International Conference on Pattern Recognition (ICPR'06), 2006. 10.1109/ICPR.2006.479.
- Ng, H., Ong, S., Foong, K., Goh, P.-S., & Nowinski, W. (2006). Medical image segmentation using k-means clustering and improved watershed algorithm. In *2006 IEEE southwest symposium on image analysis and interpretation (ICPR'06)*, IEEE, 3, (pp. 850–855). 10.1109/SSIA.2006.1633722.
- Novello, S., & Le Chevalier, T. (2003). Chemotherapy for non-small-cell lung cancer. Part 1: Early-stage disease. *Oncology*, 17(3), 357–364.
- Orsatti, G., Morosi, C., Giraudo, C., Varotto, A., Crimi, F., Bonzini, M., Minotti, M., Frigo, A.C., Zanetti, I., & Chiaravalli, S. (2020). Pediatric rhabdomyosarcomas: Three-dimensional radiological assessments after induction chemotherapy predict survival better than one-dimensional and two-dimensional measurements. *Cancers*, 12(12), 3808. doi.org/10.3390/cancers12123808.
- Osher, S., & Sethian, J. A. (1988). Front propagation with curvature dependent speed: Algorithms based on Hamilton. *Journal of Computational Physics*, 79(1), 12–49. [https://doi.org/10.1016/0021-9998\(88\)90002-2](https://doi.org/10.1016/0021-9998(88)90002-2)
- Quantitative Imaging Network (QIN). (2021). Lung CT. <https://wiki.cancerimagingarchive.net/display/Public/QIN+LUNG+CT> (Accessed: June 2021).
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779–788).
- Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
- Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. In *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7263–7271).
- Rezatofighi, H., Tsai, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. In *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 658–666).
- Sang, J., Alam, M.S., & Xiang, H. (2019). Automated detection and classification for early stage lung cancer on CT images using deep learning. In *Pattern Recognition and Tracking XXX, International Society for Optics and Photonics, SPIE*, 10995, (pp. 200–207). 10.1117/12.2520333.
- Shrivastava, K., Gupta, N., & Sharma, N. (2014). Medical image segmentation using modified k means clustering. *International Journal of Computers and Applications*, 103 (16), 12–16.
- Siegel, R. L., Miller, K. D., & Jemal, A. (2019). Cancer statistics. CA: A Cancer Journal for Clinicians, 69(1), 7–34. <https://doi.org/10.3322/caac.21551>
- Sohaib, S., Turner, B., Hanson, J., Farquharson, M., Oliver, R., & Reznek, R. (2000). CT assessment of tumour response to treatment: Comparison of linear, cross-sectional and volumetric measures of tumour size. *British Journal of Radiology. Supplement*, 73 (875), 1178–1184. <https://doi.org/10.1259/bjr.73.875.11144795>
- Sori, W. J., Feng, J., & Liu, S. (2019). Multi-path convolutional neural network for lung cancer detection. *Multidimensional Systems and Signal Processing*, 30(4), 1749–1768.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1), 1929–1958.
- Sun, T., Gabbouj, M., & Neuvo, Y. (1994). Center weighted median filters: Some properties and their applications in image processing. *Signal Processing*, 35(3), 213–229. [https://doi.org/10.1016/0165-1684\(94\)90212-7](https://doi.org/10.1016/0165-1684(94)90212-7)
- World Health Organization. (2021). Cancer. <https://www.who.int/news-room/fact-sheets/detail/cancer> (Accessed: August 2021).
- Xu, K., Jiang, H., & Tang, W. (2020). A New Object Detection Algorithm Based on YOLOv3 for Lung Nodules. In *In Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence* (pp. 233–239). <https://doi.org/10.1145/3404555.3404609>
- Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., & Yoo, Y. (2019). Cutmix: Regularization strategy to train strong classifiers with localizable features. In *In Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 6023–6032).
- Zhang, C., Sun, X., Dang, K., Li, K., Guo, X. W., Chang, J., ... Liang, Z. (2019). Toward an expert level of lung cancer detection and classification using a deep convolutional neural network. *The Oncologist*, 24(9), 1159–1165. <https://doi.org/10.1634/theoncologist.2018-0908>
- Zhang, L., Li, Y., Chen, H., Wu, W., Chen, K., & Wang, S. (2021). Anchor-free YOLOv3 for mass detection in mammogram. *Expert Systems with Applications*, 191, Article 116273. <https://doi.org/10.1016/j.eswa.2021.116273>
- Zhu, L., & Spachos, P. (2021). Support vector machine and YOLO for mobile food grading. *Internet of Things*, 13, Article 100359. <https://doi.org/10.1016/j.iot.2021.100359>
- Zhu, R., & Wang, Y. (2012). Application of improved median filter on image processing. *Journal of Computers*, 7(4), 838–841. <https://doi.org/10.4304/jcp.7.4.838-841>