



Image segmentation review: Theoretical background and recent advances

Khushmeen Kaur Brar^a, Bhawna Goyal^{a,b,*}, Ayush Dogra^c, Mohammed Ahmed Mustafa^d, Rana Majumdar^e, Ahmed Alkhayyat^f, Vinay Kukreja^c

^a Department of ECE and UCRD, Chandigarh University, 140413, Mohali, Punjab, India

^b Faculty of Engineering, Sohar University, Sohar, Oman

^c Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

^d Department of Medical Laboratory Technology, University of Imam Jaafar AL-Sadiq, Iraq

^e School of Engineering, Department of Computer Science, Sister Nivedita University, Kolkata, India

^f College of Technical Engineering, The Islamic University, Najaf, Iraq



ARTICLE INFO

Keywords:

Image segmentation
thresholding
deep learning (DL)
semantic segmentation
convolutional neural network (CNN)
medical image segmentation
segmentation evaluation
research
technology

ABSTRACT

Image segmentation is a significant topic in image refining and automated image analysis with relevance for instance object recognition, diagnostic imaging scanning, mechanized perception, monitoring cameras, satellite imaging, and image compression, and so on. This technology has become an essential component of image assessment as it facilitates the depiction, taxonomy, and conception of the subject matter in the representation. The latest advances in computer vision procedures and the progressive attainability of substantial databases have made it absolutely typical in the computer vision domain. Lately, because of the progression of deep learning techniques, it is observed that a considerable number of tasks are directed at establishing image segmentation strategies operating deep learning models. As an evolving biomedical image refining mechanization, medical image segmentation has computed significant improvements to sustainable health maintenance. Presently it has evolved into a predominant experimentation direction in the domain of computer vision. With the rapid evolution of deep learning, diagnostic image scanning characterized by deep convolutional neural networks has become a research epicentre. This review covers a survey on existing image segmentation approaches into extensive categorization of their algorithms. Additionally, this review outlines the therapeutic and non-therapeutic image databases deployed in the literature for implementing the experimentation. Apart from this, numerous evaluation metrics are discussed for evaluation comparing the results of different segmentation techniques. Further, a detailed discussion on the distinct domains of applications in image segmentation is provided. In conclusion, a discussion on several issues, especially in therapeutic domain and scope in the domain of image segmentation for implementation in the diverse disciplines is provided

1. Introduction

Image segmentation is the most prevalent research domain in deep learning, and provides the foundation of pattern recognition and image understanding. The expansion of image segmentation methodologies is associated with numerous fields of study such as automated cars, smart therapeutic technology [1], image search engines, industrial inspection, and augmented reality. Technological innovation on the image segmentation process has accomplished immense advancement both in hypothesis and implementation. It is a crucial section of image refining. It has emerged as an epicentre in the domain of image recognition. Image segmentation partitions the whole image into various parts that

contain some equivalent properties. Lately, such approaches are augmenting in a rapid and more explicit direction [2,3]. Image segmentation finds application in pattern recognition and image categorization in distinct domains like farming [4], medical [5,6,7], and forensic [8,9]. Various analysts investigated that the segmentation of pictures is most critical stage for refining as the quality of such pictures transformed the outcomes of the remaining procedures. Nevertheless, the segmentation technique has emerged a complex matter because of the complicated background and inconstant radiance on the pictures. Poor quality of segmented pictures might induce erroneous and inefficacious result. The standard segmentation approaches generally address emphasizing and attaining the details present in a single image that usually involves

* Corresponding author.

E-mail address: bhawnagoyal28@gmail.com (B. Goyal).

professional expertise and human mediation. But it is critical to achieve high-level semantic information from images. Co-segmentation techniques incorporate recognizing typical entities out of a series of images that involves the acquisition of specific preliminary knowledge. As the image observation of these procedures is superfluous, they are categorized as semi-supervised or weakly supervised approaches. With the enhancement of exhaustive fine-grained interpretation image datasets, image segmentation approaches derived from deep neural networks (DNN) have progressively turned into a prevalent field. Even though numerous implementations have been modelled in image segmentation analysis, there are yet many hassles like attribute depiction, model arrangement, and optimization. Specifically semantic segmentation is yet entirely challenging because of insufficient or restricted notations, class variance, over-fitting, extensive training time, and gradient withdrawal. One of the key objectives of computer vision specialists is to increase segmentation accuracy. For many applications, such as self-driving vehicles, object recognition, and medical imaging, accurate segmentation is crucial. Even while segmentation accuracy has significantly improved recently because of deep learning approaches, much more can be done. In order to increase segmentation accuracy, researchers are attempting the following methods:

- Including other data sources: Adding data sources other than the raw picture data, is a way to increase the accuracy of segmentation. For instance, depth data can be a useful indicator for segmenting objects and determining their boundaries, especially in intricate situations with clutter and occlusions.
- Creating better segmentation algorithms: Scientists are always creating better, more accurate segmentation algorithms for images. To improve segmentation outcomes, for instance, some recent methods employ reinforcement learning or adversarial training.
- Enhancing annotation quality: Obtaining high accuracy in segmentation algorithms depends on the caliber of the ground truth annotations employed for training. Scholars are employing diverse techniques to enhance the quality of annotations, such as integrating domain expertise and leveraging crowdsourcing portals.

4. Improving assessment metrics: Rating the precision of segmentation algorithms is largely dependent on evaluation measures. Beyond the conventional Dice coefficient and Jaccard index, investigators are investigating novel metrics for assessment that can more accurately reflect the quality of object borders, such as the Boundary F1 score.

Combining Deep Learning with Conventional Methods:

Even though deep learning methods have demonstrated impressive results in segmentation tasks, conventional methods like thresholding, clustering, and morphological procedures can still yield insightful findings and increase precision.

Researchers are combining deep learning with conventional methods for picture segmentation in the following ways:

- Hybrid models: Scholars are creating approaches that incorporate both conventional methods and deep learning. Certain methods, for instance, establish deep learning models or post-process segmentation data using clustering or thresholding.
- Multi-stage strategies: These strategies start with segmentation utilizing deep learning and refine the outcomes using conventional methods. For instance, certain methods smooth and improve segmentation results by using morphological procedures.
- Attention-based techniques: These algorithms are a class of deep learning models that use conventional methods to determine the weights assigned to each feature in a feature map. By concentrating on pertinent visual elements and disregarding unnecessary ones, attention-based approaches can increase accuracy.
- Transfer learning: It is the process of fine-tuning deep learning algorithms for particular segmentation tasks after pretraining them on

huge datasets. Features that are appropriate for transfer learning can be found using conventional methods like thresholding or clustering.

Applications of image segmentation:

Numerous industries, such as medical imaging, robotics, self-driving vehicles, and monitoring use image segmentation extensively. Here are a few instances of several fields where image segmentation is applied:

Therapeutic imaging: For purposes like tumor identification, tissue segmentation, and illness evaluation, image segmentation is frequently employed in healthcare imaging. Defining a course of therapy and keeping track of the progression of an illness require accurate segmentation.

Robotics: It uses image segmentation to recognize and manipulate objects. For instance, in manufacturing plants, segmentation can be used by robots to identify and grab particular things, like devices or parts.

Self-Driving vehicles: The ability of self-driven vehicles to recognize and categorize things in their surroundings, including other cars, pedestrians, and barriers, depends on image segmentation. Precise division is essential for secure and dependable self-navigating.

Image segmentation has applications in monitoring to follow and identify individuals and objects in real-time video streams. Classifying and identifying objects of interest, like unusual conduct or potential dangers can be aided by segmentation.

Farming: Crop inspection, disease diagnosis, and estimation of yield all require image segmentation. Farmers may maximize the production of crops by making better choices about handling crops with the aid of accurate segmentation.

Creative Pursuits: In the fields of art and design, picture segmentation is utilized for activities including correcting colors, style transfer, and image manipulation. Using segmentation can make it easier to apply various effects or alterations to different items or sections within a picture.

The main contributions in this research are as follows:

- This review incorporates more than over 300 relevant papers to scrutinize discrete image segmentation approaches advocated by the analysts in the intervening years.
- An extensive overview on different segmentation approaches according to distinct categories is submitted.
- The review addresses different statistical assessment metrics and substantial benchmark datasets employed in the literature to assess the execution of diverse approaches in the domain of image segmentation.
- For analysing the execution of image segmentation frameworks, experimental estimates have been executed employing different datasets.
- The paper takes account of the recent literature proportionate to segmentation issues, and outlines numerous procedures classified into distinct classes.
- We present an extensive survey and evaluation of divergent phases of segmentation frameworks utilizing DL.

The rest of this paper is structured as follows: Section II demonstrates the overview of image segmentation frameworks predicated on the primitive philosophy and the elaborate interpretation on distinct classifications of techniques. Section III includes image segmentation-based datasets for 2 D images. Section IV presents the details of distinct evaluation metrics utilized for evaluating the segmentation results. Section V comprises experimental evaluation on medical images and other 2 D images to evaluate the implementation of distinct algorithms. Section VI presents diverse domains of applications of image segmentation. The conclusion based on interpretation and scope is provided in Section VII.

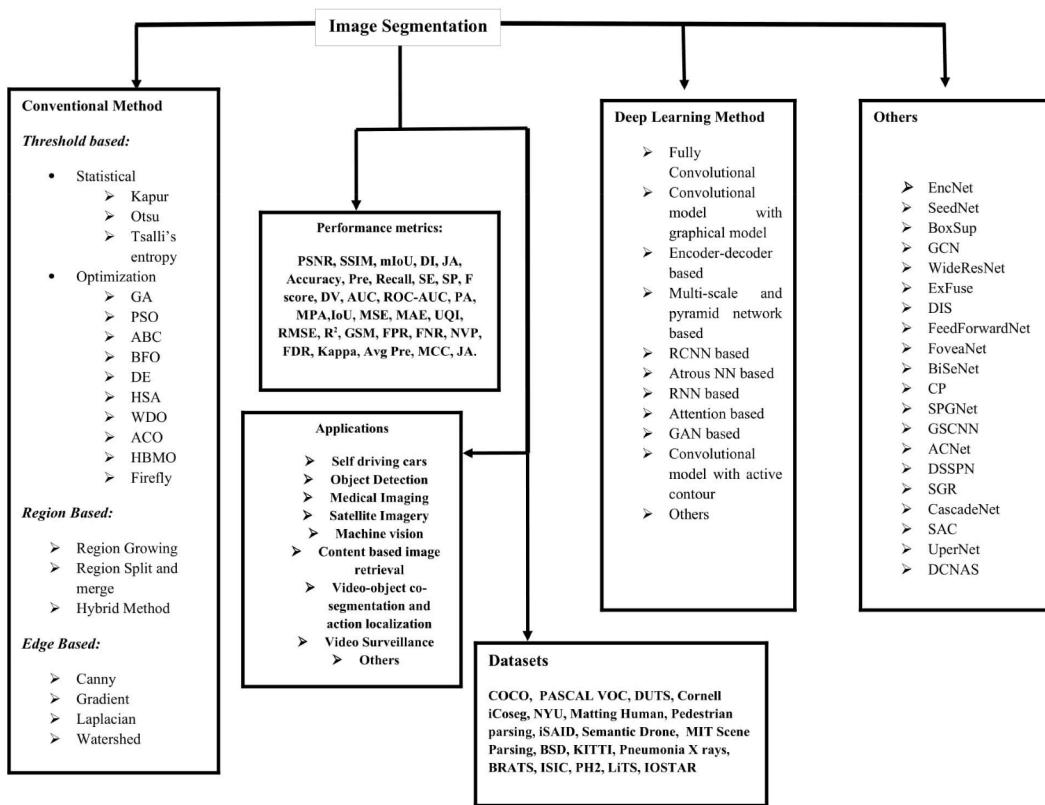


Fig. 1. Taxonomy of image segmentation approaches

2. Overview and background concepts

This section illustrates regarding the enlightenment and approaches that involve acknowledging the research work. The fundamentals of threshold based, region based, edge-based, watershed based and cluster based, their architecture, are exhibited here. In this section the significant explanation and relevance of these types is reported and illustrated in Fig. 1.

2.1. Conventional methods

2.1.1. Thresholding based segmentation technique

It is one of the most uncomplicated techniques for segmenting pictures. It is a typical region segmentation approach [10,11], where the picture is partitioned to sets of pixels having values either lesser, or higher than or equivalent to the threshold. Numerous thresholding approaches have been evolved [12], like global techniques originated from grey-level histograms, global technique originated from local features, and adaptative thresholding.

The global threshold presume that the image contains a dual-mode histogram, demonstrating a deep valley enclosed by two discrete ranges, where one of the ranges indicates the entities, while the other indicates the background of the image. The two modes to achieve image thresholding are bi-level and multilevel thresholding. The former approach segments the picture to two categories: background and foreground. This technique produces binary image as a resultant segmentation, and is therefore not appropriate for grayscale and RGB images. On the contrary, the latter one picks numerous threshold values to partition the image into distinct categories associated with various entities and background. This technique is preferred and extensively utilized in the segmentation architecture because it exhibits more details by means of the derived segmented images. In spite of that, as the segmentation proportion escalates in multilevel thresholding problem, estimating complications escalate too [13].

In general, there are two outlooks for multilevel image segmentation: traditional method [14] and meta-heuristic method [15]. The multilevel image thresholding manoeuvring traditional approaches is inefficient because it explores the finest values to revise the choice criterion; conversely meta-heuristic methods are computationally systematic and consequently are attaining much preference in interpreting the image segmentation complications. Of all the phenomenal thresholding-based segmentation strategies, entropy-based approach has inspired various analysts. The principal issue amid picture thresholding-based strategies is the choice of suitable threshold term that can precisely differentiate entities from background and other units prevailing in the picture. Several types of numerical features have formerly been presented for picking a threshold term in accordance with image histogram like maximum likelihood [16], entropy [17] and class variance [18]. Over the past several decades, the utilization of meta-heuristic algorithms has been proliferating amid the most common segmentation strategies. Frequent ways for thresholding are presented in this paper to partition the picture and withdraw significant targeted specifications [19,20]. Thresholding categorizes the grayscale or RGB image pixels into divisions derived from their intensity level (l). To properly determine the divisions in an image, the optimal threshold value (T) should be fixed as demonstrated by the following Eq. (1):

$$\begin{aligned} A_1 &\leftarrow m && \text{if } 0 \leq m \leq T \\ A_2 &\leftarrow m && \text{if } T \leq m \leq l - 1 \end{aligned} \quad (1)$$

where m specifies one of the x^y pixels of grayscale image I characterized in one grayscale levels $l = \{0, 1, 2, 3, L-1\}$. A_1 and A_2 signifies the two distinct categories scattered employing pixel m .

The idea of bi-level thresholding can be augmented to multilevel thresholding that deploys numerous threshold values for generating an output image with various categories using Eq. (2)



Fig. 2. (a) Original image from PH2 dataset; (b) Gray scaled image; (c) Otsu segmented image

$$\begin{aligned} A_1 &\leftarrow m & \text{if } 0 \leq m \leq T_1 \\ A_2 &\leftarrow m & \text{if } T_1 \leq m \leq T_2 \\ A_r &\leftarrow m & \text{if } T_r \leq m \leq T_{r+1} \\ A_z &\leftarrow m & \text{if } T_z \leq m \leq l-1 \end{aligned} \quad (2)$$

where T_1, T_2, T_r, T_z represent different thresholds. The intent is to pick the T values that correctly distribute the discrete parts in an image utilizing bi-level or multilevel thresholding measures.

An examination on explicit thresholding procedures and assessment of their computable performance is presented in [21]. An evaluation for multilevel thresholding by implementing optimization application has been depicted by [22]. Similarly a review presenting the contrast evaluation of relative entropy and entropy thresholding approaches has been depicted in [15]. A review on distinct meta-heuristic strategies and their contrast has been presented in [23]. For procuring prime multilevel threshold, numerous heuristic maximization systems have been administered to unravel the issues in reference to multilevel thresholding. In the course of time, in the literature, diverse projects accordant with swarm algorithms like genetic algorithm (GA) [24,25,26] imparts a typical system architecture to resolve complicated maximization issues. Differential evolution (DE) [27,28,29] is certainly an impressive information retrieval system for unravelling solitary objective maximization issues. Ant colony optimization (ACO) is another prevalent swarm that has been exercised to broad range of maximization problem. It is actuated by cooperative demeanour of ants that are performing complicated chores with no leader [30]. The *E. coli* bacteria found in the intestine exercises a foraging strategy that motivated analysts to exploit it as an optimization strategy known as bacterial foraging optimization (BFO) [31,32,33]. An innovative approach motivated from the composers improvising new melodiousness while playing led to harmony search algorithm (HSA) [34]. A comprehensive optimization approach which imitates the electromagnetism law of physics promotes electromagnetism optimization [35]. The algorithm motivated by the means of breeding in honey bees give rise to honey bee mating optimization (HBMO) [36,37]. Similarly another approach related to this category is firefly algorithm [38,39,40]. A distinct approach utilized to mimic the hunting behaviour of honey bee colonies is artificial bee colony [41,42,43,44,45,46,47,48]. PSO [49,50,51,52] is also an innovative maximization approach formulated on swarm intelligence. One more algorithm which is motivated by the movement of wind is wind-driven optimization (WDO) [53,54] have been implemented to emphasize several multilevel picture segmentation liabilities for estimation of optimal threshold.

2.1.2. Statistical based segmentation

In time domain mechanism, grey-level scattering of each class incorporates a probability density function (PDF), which is frequently anticipated to observe the Gaussian distribution. These mediums of time-domain tactics aim to approximate the variables of scatter that will best fit the specified histogram features. It consistently accelerates an inconsistent optimization issue that has a computationally expensive result [23]. A thresholding approach that reduces the categorizing inaccuracy probability by resembling the histograms deploying the fusion of conventional distributions has been discussed in [55]. Similarly a fusion optimization measure to accomplish thresholding employing

Gaussian function fitting and Otsu's class variance has been discussed in [56]. A fusion optimal computation means for image segmentation is reported in [57]. Statistical-based approach picks the threshold to partition the grey-pixel regions of the image in an excellent plan by means of numerous discriminatory measures.

2.1.2.1. Kapur's thresholding method. This technique enhances the entropy of segmented histogram to a degree which every differentiated part comprise more consolidated spread [31]. It is formed on the probability spread of grey-level histogram. Provided, the entity and its background are carefully separable from one another, the grey-level histogram is categorized so as to pick the thresholds at the extremity of the valley [22,42,43]. The method can be well explained by the equation 3:

$$\begin{aligned} H &= - \sum_{r=1}^{T_1-1} \frac{m_r}{\omega_0} \ln \frac{m_r}{\omega_0}, \omega_0 = \sum_{r=0}^{T_1-1} m_r \\ H_a &= - \sum_{r=T_a}^{T_{a+1}-1} \frac{m_r}{\omega_a} \ln \frac{m_r}{\omega_a}, \omega_a = \sum_{r=T_a}^{T_{a+1}-1} m_r \\ H_b &= - \sum_{r=T_b}^{l-1} \frac{m_r}{\omega_b} \ln \frac{m_r}{\omega_b}, \omega_b = \sum_{r=T_b}^{l-1} m_r \end{aligned} \quad (3)$$

The complication of multilevel thresholding is composed as a p -dimensional maximization issue. To deal with the complication of multilevel segmentation of a picture, p spatial maximization measures of threshold are noted for intensification of the decision criterion through Eq. (4):

$$\overrightarrow{(T)} = \operatorname{argmax} \left(\sum_{r=0}^p H_a \right) \quad (4)$$

2.1.2.2. Otsu thresholding. The approach suggested by Otsu is a technique for ascertaining ideal thresholds. Otsu relies upon the DE to enhance dissociable measure of the groups in grey-scale image [58]. The basic Otsu thresholding outcomes for segmenting melanoma is illustrated in fig. 2. Otsu generated the inter-class dissimilarity as total of sigma values of every class demonstrated in Eqs. (5–6):

$$f(t) = \sigma_1 + \sigma_2 \quad (5)$$

$$\sigma_1 = \omega_1(\mu_1 - \mu_T)^2, \sigma_2 = \omega_2(\mu_2 - \mu_T)^2 \quad (6)$$

where μ_T signifies the average magnitude of input picture. The mean level i.e., μ_i of two labels for the contingency of bi-level thresholding is attained by utilizing Eq. (7) [59].

$$t' = \operatorname{argmax}(f(t)) \quad (7)$$

Besides, inter-class dissimilarity can be enhanced for issues of multilevel thresholding as illustrated in Eq. (8):

$$f(t) = \sum_{i=0}^n \sigma_i \quad (8)$$

The sigma value is stimulated by deploying Eq. (9), and the mean

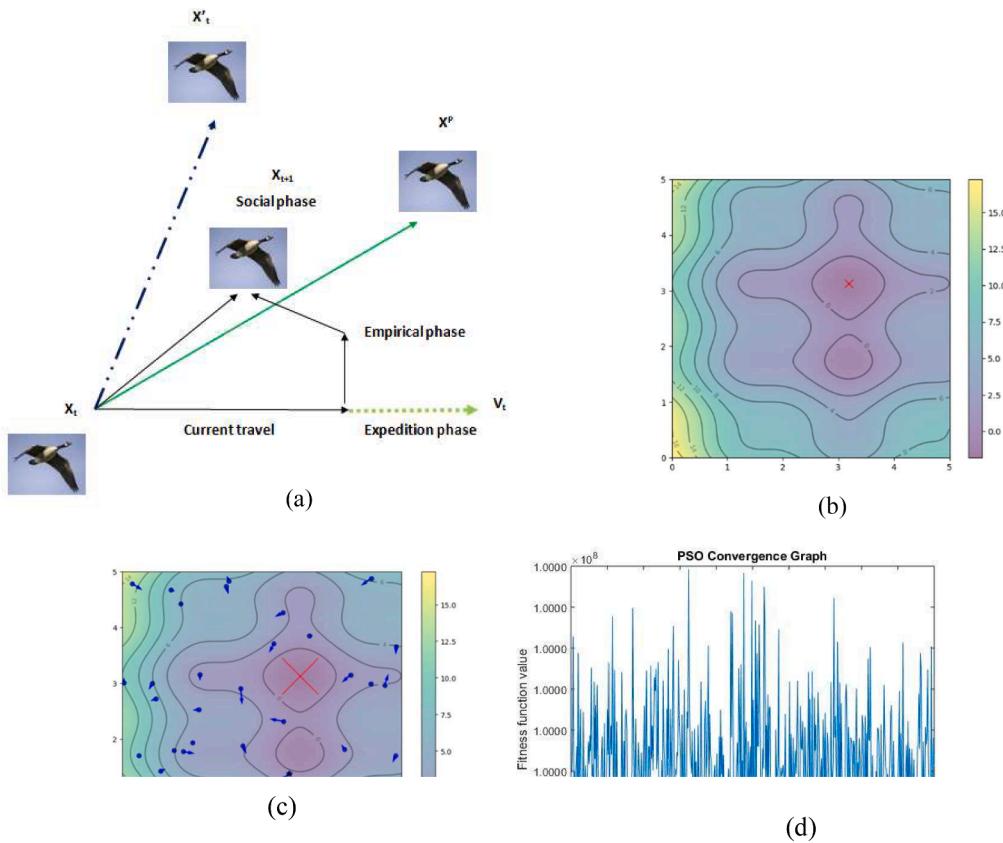


Fig. 3. (a) PSO algorithm; (b) Updated position of current particle; (c) Score after 1st iteration; d)PSO convergence graph [70]

levels are derived utilizing Eq. (10):

$$\begin{aligned}\sigma_1 &= \omega_1(\mu_1 - \mu_T)^2, \quad \sigma_2 = \omega_2(\mu_2 - \mu_T)^2, \quad \sigma_k = \omega_k(\mu_k - \mu_T)^2, \\ \sigma_z &= \omega_z(\mu_z - \mu_T)^2\end{aligned}\quad (9)$$

$$\mu_1 = \sum_{i=1}^{t_1-1} \frac{im_i}{\omega_i}, \quad \mu_2 = \sum_{i=t_2}^{t_2-1} \frac{im_i}{\omega_i}, \quad \mu_k = \sum_{i=t_k}^{t_{k+1}-1} \frac{im_i}{\omega_i}, \quad \mu_z = \sum_{i=t_z}^{L-1} \frac{im_i}{\omega_i}, \quad (10)$$

The ideal multilevel thresholding is deduced using choice criterion maximization as depicted by Eq. (11):

$$\left(\vec{t} \right) = \operatorname{argmax} \left(\sum_{i=1}^z \sigma_i \right) \quad (11)$$

2.1.2.3. Tsalli's entropy-based thresholding method. The technique evaluates the comprehensive and equitable characteristics of the histogram that can be certainly administered for multilevel thresholding by intensifying the Tsalli's entropy. The entropy for diverse architect is attained exploiting PDF, where $P = \{P_a\}$ signifies the possibility of obtaining the framework in every "a" feasible condition. $0 \leq P_a \leq 1$ and $\sum_{a=0}^z P_a = 1$

$$S = - \sum_{a=1}^f P_a \ln(P_a) \text{ (Shannon property)} \quad (12)$$

According to multifractal theory, it is summarized into non-substantial arrangement implementing entropic code as in equation 13 [60]:

$$S_v = \frac{1 - \sum_{a=1}^f (P_a)^v}{v-1} \quad (13)$$

where f is the overall likelihood of the procedure and v is a unit that computes the intensity of non-substantiality of the scrutinized

framework designated entropic index with the following Eq. (14):

$$S_v(A+B) = S_v(A) + S_v(B) + (1-v).S_v(A).S_v(B) \quad (14)$$

This type of mode is utilized for thresholding of the image. Let's assume in an input image, R grey levels exist ranging $\{1, 2, \dots, R\}$, and $\{P_a = P_1, P_2, \dots, P_R\}$ signifies probability spread of the image gray-level magnitude locales. Referring to this probability spread, 2 categories i.e., category A for background and category B for scrutinized entity, are depicted by equation 15:

$$P_A = \frac{P_1}{P^A}, \quad \frac{P_2}{P^A}, \dots, \frac{P_t}{P^A}, \text{ and } P_B = \frac{P_{t+1}}{P^B}, \quad \frac{P_{t+2}}{P^B}, \dots, \frac{P_R}{P^B} \quad (15)$$

$$\text{Here } P_A = \sum_{a=1}^t P_a, \text{ and } P_B = \sum_{a=t+1}^R P_a.$$

The Tsalli's entropy can also be depicted by the Eq. (16):

$$S_v^A(t) = \frac{1 - \sum_{a=1}^t (P_a/P^A)^v}{v-1}, \quad S_v^B(t) = \frac{1 - \sum_{a=t+1}^R (P_a/P^B)^v}{v-1} \quad (16)$$

When the calculation of data for both the categories A and B is boosted, only corresponding grey level is titled as an ultimate threshold value: $T' = \operatorname{argmax} [S_v^A + S_v^B + (1-v).S_v^A(t).S_v^B(t)]$ (17)

2.1.3. Optimization based thresholding approaches

In an optimization approach, to realize the ideal implementation, particular conditions are primarily necessary to be regulated perfectly. These optimization strategies are categorized as transformative procedures and swarm intelligence procedure. Transformative ones are naturally accommodated populations to the conditions. Such methods imitate the biological evolution and genetic construct phenomena. Getting influence from the intuitive life, analysts evolved several optimization computations for explaining the complicated issues that occurred in the past several decades.

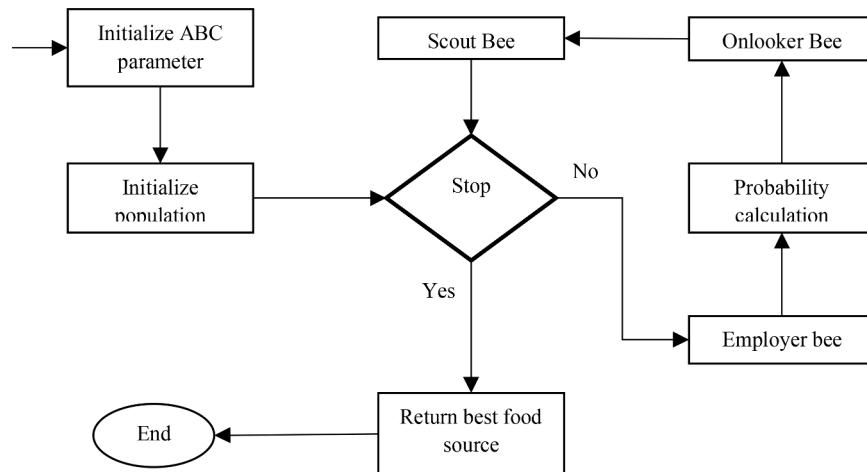


Fig. 4. Natural selection approach for ABC methodology [74]

2.1.3.1. GA based thresholding approaches. The approach is established on the natural miscellany theories. It is an effective unbiased maximization strategy with a substantial solution space. It is a comprehensive investigating technique that imparts robust forage beyond confining into limited ideal results [61]. GA executes the lateral search; so, the choice of ideal thresholds is instant in multilevel thresholding situation. Consequently, it is favoured by various analysts to rectify the specifications of distinct segmentation of methodologies and to generate novel proficiencies [62]. The image segmentation employing 3-level thresholding referring to fuzzy segmentation, entropy hypothesis and probability division is presented in [63]. This technique has introduced the novel fuzzy entropy manoeuvring probability evaluation. Hammouche et al. developed a computerized multilevel thresholding approach regulated by GA to minimize the evaluation time for image segmentation [25]. Further to facilitate the multilevel threshold enhancement and to augment the ideal constancy efficiently, a learning operator (GA-L) is utilized in [64]. Kezong et al. proffered a novel procedure on the basis of GA for fast threshold preference computation in multilevel minimum cross-entropy thresholding (MCET) [26]. In another study, analysts have proposed a robust transformation technique of angular rotation and a joint investigation tactic to quantum GA [24]. The presentation of the multilevel thresholding of therapeutic brain pictures employing an innovative substantial encoded GA with simulated binary crossover is another approach [65]. The fusion of the GA and feed forward neural network for segmenting the pictures is illustrated in [66].

2.1.3.2. PSO based thresholding approach. PSO is a community based haphazard enhancement algorithm, where the probable conclusion is designated as particle position and the discovery scope as swarm. Because of the accelerating concurrence rate, PSO operates better than GA and ACO [67]. The disadvantage of the GA is that as the concurrent rate escalates, its ideal constancy reduces. An innovative QPSO approach contemplated to maintain the fast concurrence pace of PSO and for minimizing the computation cost and to suppress the trouble of magnitude, the cooperative QPSO or CQPSO was proposed by Gao [52]. In another finding, a novel PSO based thresholding technique is proposed that optimizes the Otsu and Kapur's decision criterion and contrasts the result with GA. The outcomes revealed that PSO has lower evaluation time than GA and Kapur's and smaller standard deviation than Otsu [68]. A learning system established PSO (LPSO) approach is recommended by authors in [69].

The better explanation of PSO algorithm is provided in [fig. 3](#). Conventionally, the PSO technique depicts the population by X, and this population is spawned employing an arbitrary number generator as mentioned by the [Eq. \(18\)](#):

$$X = L + \text{rand} \times (U - L) \quad (18)$$

where L and U depict the lower and upper limits of the search domain. Every particle of X can upgrade its velocity V by the [Eq. \(19\)](#):

$$v_i(t+1) = w v_i(t) + c_1 r_1(x_{pi}(t) - x_i(t)) + c_2 r_2(x_{gi}(t) - x_i(t)) \quad (19)$$

Cause the particle to come back to a former location better than the existing
 Cause the particle to track the leading neighbour direction
Inertia
 Cause the particle proceed in coordination with uniform velocity

where c_1, c_2 represents coefficients of acceleration, t represents time and r_1, r_2 are arbitrary values.

2.1.3.3. ABC based thresholding approach. In this procedure, location of particular food source is the probable interpretation to the specific maximization challenge. The bees are listed under managers, spectators and patrol bees. After arbitrarily producing the food sources, primarily horde are consistently dispersed as manager bees and spectator bees which are assigned for examining food spots and to ascertain the beebread volume (decision criterion) of those food spots while patrol bees disperse arbitrarily across the region. The flowchart for ABC algorithm is illustrated in [fig. \(4\)](#). The optimum Tsallis's entropy using ABC strategy to explore an ideal layered thresholding is presented by Zhang [71]. Similarly, an innovative optimized entropy-based ABC thresholding approach was presented and contrasted with four other thresholding strategies [72]. The ABC is also exploited in the exertion of SAR picture partitioning [47] where it explores the ideal threshold term in a constant grayscale span. Sushil et al., proposed a bi-level thresholding mode employing PSO and ABC assisted with Otsu methodology [48]. The importance of ABC algorithm in picking the threshold for

partitioning the picture by hypothesizing a Gaussian mixture representation, the variables of which are prosecuted by the ABC procedure is discussed in [45]. A comparison based study utilizing the ABC algorithm that carried out comprehensive analysis with the implications attained by the hypothetical methodology is discussed by Ming [73]. Akay designed 2 promising swarm-based maximization procedures, PSO and ABC statistically advocated by Kapur's approach and inter-class dissimilarity for deriving the layered thresholds. The attained outcomes reveal that the ABC approach besides entropy measure and inter-class dissimilarity can be efficiently utilized for layered thresholding. However, it is observed that both PSO and ABC approaches are a replacement for Otsu approach if the threshold value is set 2. Empirical evaluations demonstrate that both the techniques are modular and their execution time escalates at a linear pace in terms of the extent of the problem [22].

2.1.3.4. BFO based thresholding approach. A comparative study of BFO algorithm and GA thresholding for image segmentation is presented by P.D. Sathya [75]. The evaluation of study reveals that the BFO approach operates finer than GA. For solving the issues of bi-level thresholding, the idea of utilizing BFO approach for optimizing Kapur and Otsu functions is presented by Nandita et.al., [32]. Similarly for solving the issues of estimation outlay, the idea of amended bacterial foraging (ABF) approach for brain illustrations was demonstrated by the same authors [76]. In yet another paper the author presented an advanced histogram-based thresholding employing modified BFO (MBFO) algorithm. The study depicts that the overall foraging facility and the concurrent rate of the BFO algorithm can be refined, provided the optimal bacteria are drifted to the forthcoming procreations amid the complete chemospheric stages. The ideal threshold is ascertained by employing the MBF procedure for optimizing Kapur and Otsu techniques [77].

An adaptive BFO (ABFO) procedure for maximization of fuzzy entropy directed for gray picture partitioning is discussed in [32]. The mechanism that is employed in this study is a modified classical BFO haphazard maximization that dynamically picks the utilization and consideration stages in chemotropism of E-coli bacteria. The utilization of BFO in facial invasive projection is leading application of segmentation. The region of lips is assimilated and segmented by utilizing BFO thresholding [33]. An MBFO algorithm for carrying out the layered segmentation by optimizing the Tsalli's thresholding operation is discussed by Kezong [78]. In this mechanism, PSO is integrated into every chemospheric stage for restoring the overall scrutiny potential and expedite the concurrent rate of BFO.

2.1.3.5. DE based thresholding approaches. Kamal et al. executed a comparison survey of various meta-heuristic approaches employed for interpreting the thresholding issues and demonstrated that the DE approach is most efficacious with regard to solution quality. An innovative image segmentation strategy utilizing DE as a maximization tool is presented in [28]. The procedure loads the 1-D image histogram by composite Gaussian functions, characterizing a pixel point, and accordingly the threshold values and whose parameters are evaluated using DE. In another study, a DE algorithm is imposed in an effort to attain preferable computational performance. The comparative analysis exhibits that DE transcends 4 extensively investigated optimizers [79]. Kanjana et al. advocated a global multilevel thresholding practicing DE approach. The modified DE, a novel transformation tactic is presented to overcome the drawback of exponential enhancement of the complexity as the value of thresholds rises. An unconventional perspective for unsupervised categorization of vegetation analysis of spectroscopy satellite images is discussed [80].

2.1.4. Edge based Segmentation approach

This approach initially determines the edges and then occludes them by employing contouring algorithms [81]. It implies that the attributes

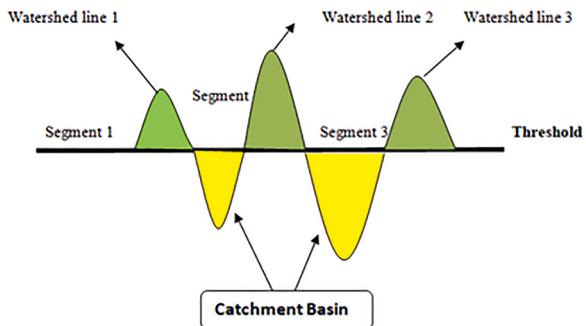


Fig. 5. Watershed based segmentation approach [116].

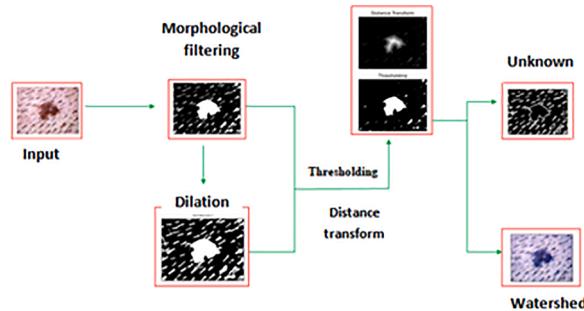


Fig. 6. . Watershed Transformation procedure for detecting melanoma.

of picture elements transform unexpectedly amid edges [82]. The edges are considered as peripheries between entities and found where transformations proceed [83]. There are several relevant algorithms for the detection of the boundaries of the entity that acquire the analytical and substantial attributes of picture entities. A range of edge sensors are employed for distinct targets. R.Jain et.al., split edge identification to 3 stages: refining, amplification, and recognition. Distinct refining techniques have been generated for inducing least blurring and reconstruction of edges [84,85,86]. Amplification emphasizes pixels where there is a considerable variation in particular intensity terms [87]. Numerous pre-programmed operators like Canny, Prewitt, Active contour are generated for detecting edges [88,89,90,91,92,93,94,95,96,97,98]. Addition to the foregoing, diverse soft computing techniques like fuzzy [99], GA [62], and Neural Network (NN)-strategies have been developed [100]. The implementation of every operator is estimated by virtue of the erroneous edge, lost edge, edge projection, and range from the accurate edge and curvature [101]. After discerning the edges, the subsequent step is to modify the margins into closed peripheries. The modification of edge frequently implies the suspension of edges which are generated by noise. Several edge-allying approaches have been developed for remedying edges which are not completely connected [102,103]. The use of Hough transform for finding the optimal edges is another tactic [104], region search to observe an entrant for connecting the pixels of edge [105]. Since numerous algorithms have been bring forward for detecting edges and concatenate them to create entities, yet determining ideal edges to produce image entities is exhausting. Additional edge-based technique depending on mathematical morphology is also efficacious, watershed transform (WT) being the widespread approach [106,107]. WT imitates authentic inundate approach and modifies the image to a slope and locate entities with a zonal facet [108, 109,110]. For this reason, WT is considered as an edge based methodology [111,112]. Furthermore, WT possesses the characteristics of both edge detection and region growing approaches [113]. The evaluation of WT extensively relies upon the algorithm prosecuted to generate the gradient that in turn induces an over-segmented outcomes because of the noise [114]. Even if numerous algorithms developed to promote the

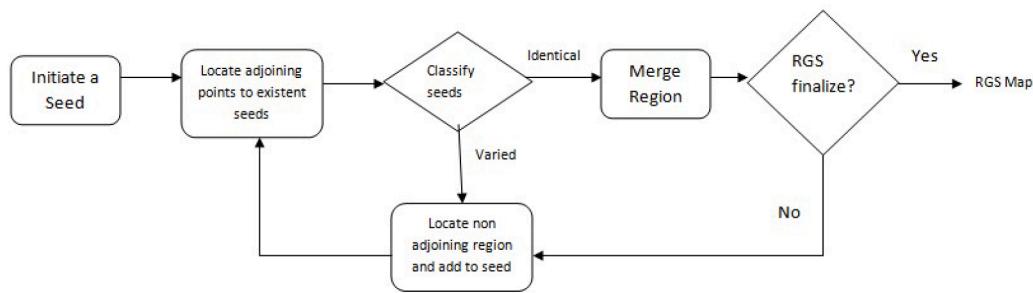


Fig. 7. Region growing and merging segmentation based approach [138]

gradient image for augmenting the effectuation of the WT, yet it creates an over-segmented picture [115]. Regardless, out of all the edge-based approaches, the WT is exhaustively practiced for image segmentation. Fig. 5 illustrates basic mechanism of watershed-based segmentation criteria [116]. Fig. 6 illustrates segmentation of skin cancer lesion using watershed approach.

2.1.5. Region based segmentation approaches

The region-based approach operates in an opposite manner than edge-based i.e., to initiate from the interior of an entity and then proliferating superficially till intersecting the margins of the entity [117]. Hypothetically, edge and region bases are dissimilar depictions of the same entity but the latter approach may prompt entirely inconsistent outcomes than edge-based segmentation [118]. Region-based techniques deduce that adjoining pixels inside the uniform region have equivalent values [119]. This technique comprises 2 fundamental steps: merging and splitting [120]. The principal steps in region-based image segmentation are: (a) to attain a preliminary (over or under) segmentation of the picture, (b) merge or split the adjoining fragments which are either identical or distinct and (c) iterating the preceding step until all the segments are grouped into merged or split classes.

2.1.5.1. Region growing and merging approach. This is a widely used and basic algorithm for region-based segmentation [121]. The limitation in this type of segmentation approach is choosing the seed point and correlation [101]. Seeded region growing contains 2 central pixel order tributaries that frame divergent resulting fragments [122]. In case, 2 or more pixels possess the similar coefficient of variance to the adjoining regions, the 1st order dependency occurs [123]. The 2nd-order dependency intervenes when a pixel possesses the similar coefficient of variance in various regions. The assignment of seed reinforces the computational cost and completion time and to solve this issue a single-seeded region growing tactic was developed [124]. A fusion

approach for segmenting the images by amalgamating the global and local region merging techniques to a coordinate frame is presented by Xueliang et al. [125]. A comparison method contingent on transformed seed growing and merging by employing a block build choice strategy has been presented by Y. Byun et al. [126]. As the subsequent step after choosing the seed is to grow the region by connecting the identical vicinal pixels, determined by a particular uniformity measure which in turn enlarges the size of the region. So, the homogeneity measure is the decisive function of confirming if a pixel have reference to the growing region or not [127]. Yu et al. [128] executed a uniformity measure incorporating the rate of the coefficient of the descent of the pixel and the weighted average of the variation among the region and the pixel. The pixels possessing low descent rates were fused to the growing region at every repetition. In such approaches, the dimensional resolution of the image has a guiding hand. The entities having lower dimensional resolution cannot be identified but the ones larger than spatial resolution will segment onto pixels. In contrast the region-merging procedure emerges from the foremost region and Multi-Resolution Segmentation (MRS) commences out of this notion [129].

An innovative Fractal Net Evolution technique (FNE), a region merging deterministic technique that initializes with 1-pixel is executed in various studies [130,131,132,133,134,135]. Besides MRS, analysts also carried out distinct region merging techniques for instance Recursive Hierarchical Segmentation (RHSeg) [136]. MS is a clustering algorithm that partitions the picture by assembling the pixels which are contiguous in the dimensional and spectral field, later adjoining the equivalent intersection points. HSWO is another clustering methodology that commences with discrete data point and continuously minimizes the units of clusters by merging. HSeg is the enhancement of HSWO in the integration procedure that includes spectrally identical but dimensionally dissociated sections in the integration stage [137]. But it necessitates too much computation time and therefore, RHSeg was developed. RHSeg splits the picture to fragments and administers HSeg

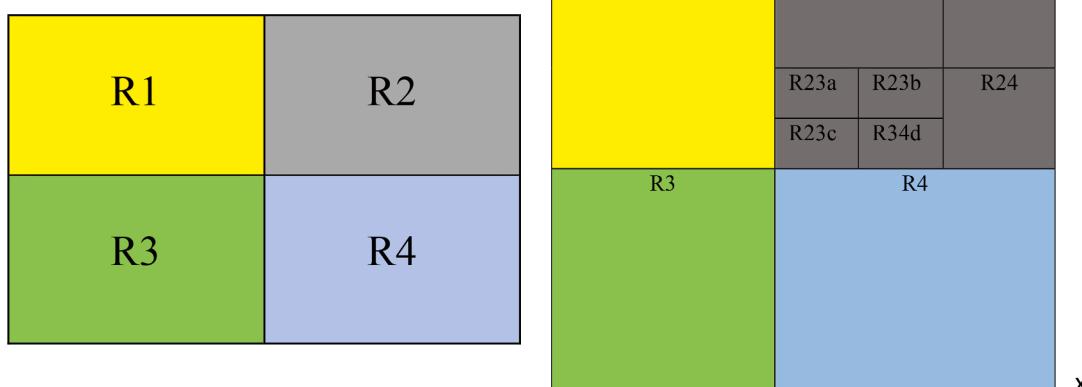


Fig. 8. (a) Illustration of original image (b). Illustration of Region split and merging segmentation (https://users.cs.cf.ac.uk/Dave.Marshall/Vision_lecture/no_de34.html)

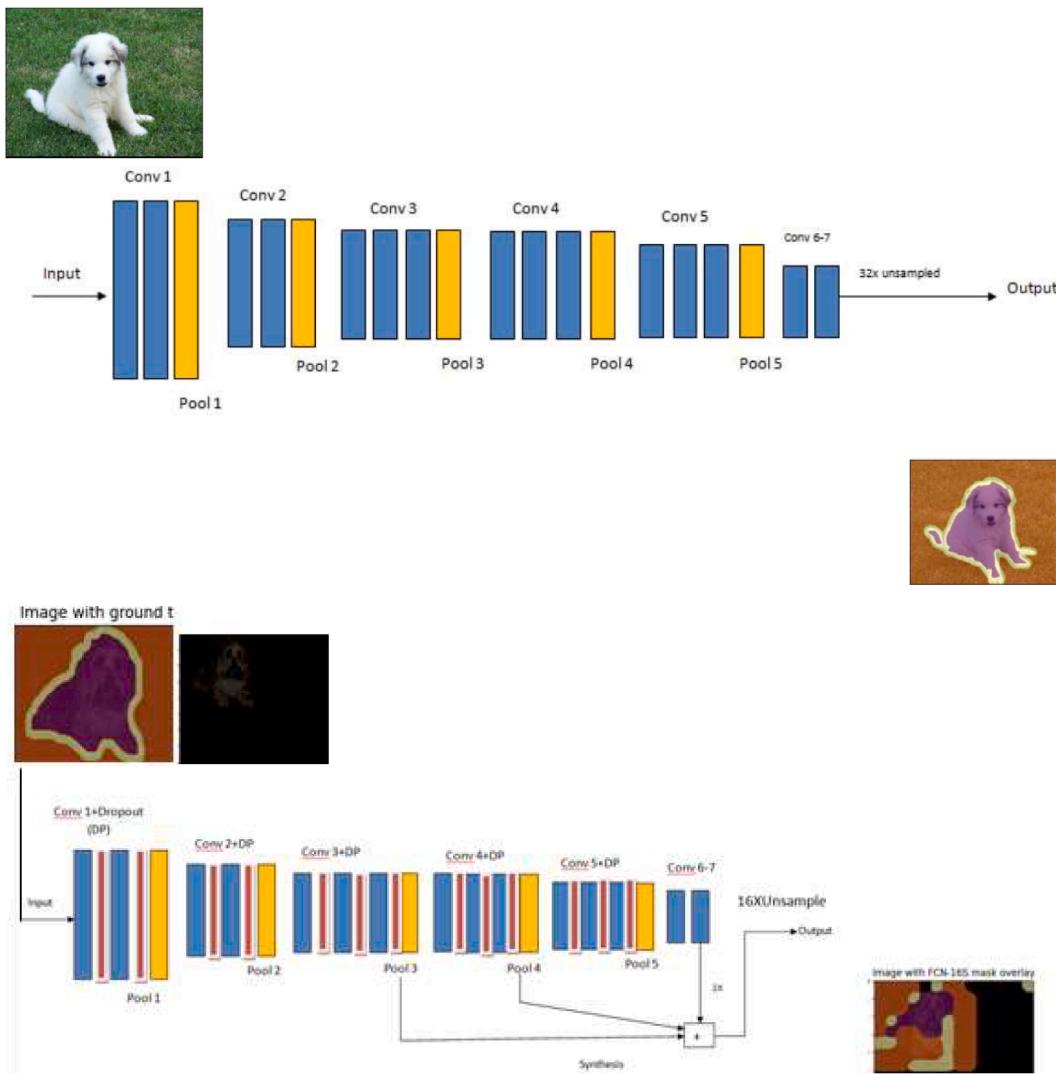


Fig .9. (a). Depiction of FCN-32 architecture [196] (b). Depiction of FCN-16 architecture [196]

on every sub-division. Ultimately, it reconnects the segmentation outcomes out of the different fragments which give rise to refining window artifacts. That means, RNSeg accommodates a supplementary step to get rid of the artifacts. The basic working of region growing/merging approach is illustrated in fig. 7.

2.1.5.2. Region splitting and merging. The split and merge was executed as a regional approach to overcome the instable execution of comprehensive measures employed in region merging and to attain accurate outcomes [112,139]. The splitting procedure initiates from the whole image, then on account of the measure for difference, partitions the image into fragments [140]. The split and merge approach fuses a bottom-up strategy with a top-down strategy [141]. The former approach creates an entity by synthesizing pixels whereas the latter approach switches from partitioning the complete image into entities based on heterogeneity specification [142]. For this method, the prime seed is simply the whole image [143]. If the seed is not uniform, the splitting approach partitions it to 4 quad sub-divisions that serve as a seed in the subsequent step and prevails till each sub-division turns uniform [83]. Deepika *et.al.* [144], presented an enhanced quad tree approach for the split-and-merge technique. I.N. Manouskas *et al.* [145] presented concepts of ersatz reinforcing and edge eradication to enhance the quality of conventional split and merge technique. The

impediment of region splitting is that the induced images imitate the data array utilized for delineating the image and emerge too square [146]. Rasha *et.al.*, [147] presented stratified merge and split segmentation subject to color, and shape attributes for road extraction from urban region pictures. Fig. 8 illustrates basic functioning of region split/merge segmentation approach.

2.1.5.3. Hybrid method (HM). For overpowering the drawbacks of the edge and region-relied techniques, analysts consolidated the conclusions of both approaches and are anticipated to bring forth preferable segmentation outcomes [148,120,149]. As already noted, edge-based approach is explicit in identifying edges, but confronted with problems in producing closed fragments. Comparatively, region-based approaches produce closed regions, but, concluded in uncertain segment peripheries [150]. Therefore, a contemporary mode in image segmentation is to implement an HM where the prime divisions are first mapped out by edge-relied technique, then integrated by region-relied approach [151,152]. Marina *et.al.* [153] fused edge and region-based approaches to bring out substantial simulated entities like agricultural fields. In the introductory portion, author draws shape details and the edge map presents an auxiliary measure in findings. In the subsequent portion, they utilized this information to regulate region growing methodology. Nan *et.al.* [154] endorsed texture clustering that was carried out as

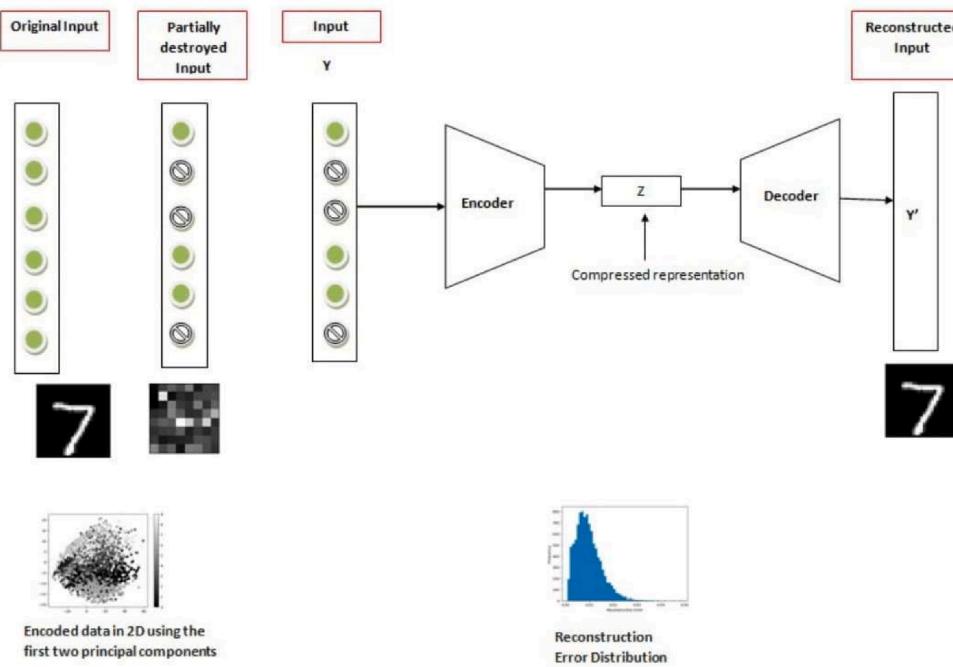


Fig. 10. The architecture of encoder-decoder model

restraints in HSWO. Merging was implemented on the basis of the region proximity and adjoining graph. Numerous region-merging techniques engage an exclusive global framework to curb the repetition procedure of merging fragments as it helps to govern under and over-segmentation. Brian [155] and Jie et al. [156] have utilized local estimates to recognize fragments which are under and over segmented at the stipulated ideal scale framework and additionally processed by relevant split and merge. Zhihua et al. [157] implemented an HM examining the idea of non-uniformity and corresponding uniformity through the merging procedure.

2.1.5.4. Semantic segmentation approaches. Machine learning (ML) has demonstrated its effectiveness in several applications over the past few years which influences remote sensing domain too. ML computations are “approximators” that learn from the training data and function in accordance. Different semantic segmentation approaches can be listed as Markov Random Field (MRF) [158,159,160,161], Bayesian Network [162], NN [163,164], active Support Vector Machine (aSVM) [165], weighted aggregation [166] and Deep Convolution Neural Network (DCNN) [167]. In semantic approach, every pixel is assigned a class label of its enclosed entity. Of all the semantic approaches, MRF elucidates a significant proportion [168,169]. It is a stochastic process that occupies the circumstantial limits across the adjoining pixels. For procuring perfect fragments out of noisy images having the compound and prominent-texture basis, an assimilated MRF model with Multi-Region Resolution (MRR) approach was presented by Chen et.al. [169]. Other innovative procedures based on semantic methodology are edge penalty function [170], discrete wavelet transform (DWT) [171], multiscale approach [172], DCNN [173,174,175] and region-based tactics with MRF to partition high-resolution images [176]. Although semantic approaches displayed promising outcomes yet they are confronted with various problems such as, in high-resolution images, it is not easy to determine appropriate attributes with semantic meaning because of excessive texture [177]. Furthermore, the scale obtained on such pictures make it complicated in ascertaining semantic standards that can distinguish entities in diverse scales [178]. Besides, semantic approach

also experience computational load to withdraw configuration information, necessitates a good quantity of training data and a large number of parameters for adjustment [179].

2.1.6. DL based approaches

In this part we provide an outline of noted DL frameworks employed by the CV network, comprising convolutional NN (CNN) [180], recurrent NN (RNNs) and long short term memory (LSTM) [181], encoder-decoders [182], and generative adversarial networks (GANs) [183]. In some instances, these approaches are trained from very beginning on novel operations but in several other instances there are not enough labelled data accessible for training a model, for which transfer learning is employed. In image segmentation process, various analysts utilized a prototype trained on ImageNet, as the encoder component, and reinforcing the prototype from fundamental weights. The hypothesis is that the reinforced models must be competent to acquire the semantic details of the image necessary for segmentation, and thus facilitating them for prototype training with limited labelled specimens. CNNs are one of the keys and extensively employed designs in the DL network, primarily for CV assignments. CNNs primarily include 3 layers: i) convolutional layers, where the majority of estimation takes place; ii) non-linear layers, that exercises an activation function on working model; and iii) pooling layers. The entities in layers are associated; which means, every entity acquires weighted information from a small proximity of entities in the prior layer. The prominent CNN configurations are: AlexNet [184], VGGNet [185], ResNet [186], GoogLeNet [187], MobileNet [188], and DenseNet [189].

2.1.6.1. FCN. The first DL method based on FCN that operates for semantic segmentation was presented by Evan et al. [190]. An FCN contains just the convolutional layer that empowers it to hold an image of random proportion and generate a segmentation layout of the identical measurement. The framework for FCN is illustrated in fig 9. For controlling the uncertain dimensions of input and output, a modified CNN framework is developed by substituting each fully-connected layer to the fully-convolutional layer. In this a dropout layer is attached

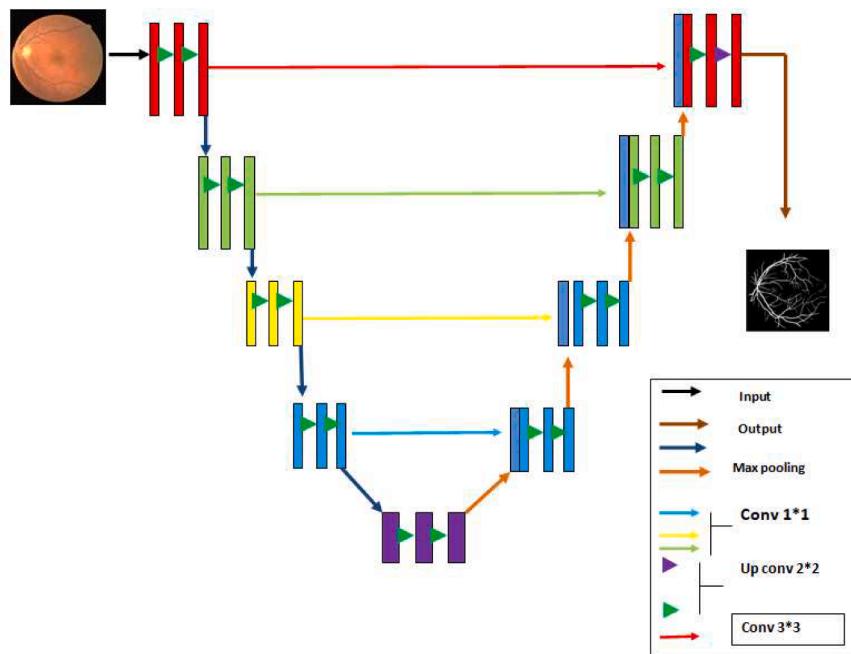


Fig. 11. Depiction of U-Net based segmentation approach [221]

following every convolutional layer for exquisite optimization. These dropout layers are typically utilized in progressive deep frameworks for exceptional coordination. The modified FCN-16 framework is shown in fig. 9(b). Consequently, the paradigm yields a dimensional segmentation layout contrary to analysis scores. By utilizing the skip connections, the architecture associates semantic data for the purpose of generating precise and elaborated segmentations. This approach is addressed as a turning point in segmentation of a picture, indicating that DL methods can be trained for semantic segmentation on variable sized data. In spite of its popularity and efficacy, the traditional FCN network possesses certain shortcomings such as it is not fairly quick for synchronous generalization, and is not conveniently negotiable to 3D data. To overcome these shortcomings, Wei et al. [191] put forward a model named ParseNet. FCNs have been implemented to diverse segmentation implementations, like brain tumor segmentation [192], instance aware segmentation [193], skin abrasion partitioning [194], and segmentation of iris [195].

2.1.6.2. Convolutional graphical approaches. Since FCN disregards crucial semantic context, distinct strategies like Conditional RF (CRF) and MRF, into DL networks are incorporated. Liang et al. [197] recommended a semantic segmentation approach derived from the amalgamation of CNNs and FC CRF. For overcoming the flawed attributions of deep CNNs, authors assembled the observations at the terminal CNN layer and a FC CRF. It was revealed that the framework was suitable for localizing segment edges with a higher efficiency than it was practical with former techniques. Alexander et al. [198] generated a FC deep structured arrangement for segmenting the picture which collectively trains CNNs and FC CRFs for segmentation, and attained promising outcomes. Shuai et al. [199] presented an identical segmentation technique by merging CRF and CNN. Guosheng et al. [200] recommended a proficient computation for segmentation established on deep CRFs by considering “patch-patch” and “patch-background” context for enhancing the segmentation. Ziwei et al. [201] introduced a semantic segmentation approach which comprises broad information toward MRFs, involving most significant connections and amalgam of label contexts.

2.1.6.3. Encoder-decoder based approaches. This is another classification of DL networks for segmenting the data. Image labeling work with encoder-decoder methodology has been extensively surveyed lately [202,203,204,205,206,207,208,209]. In the foremost interpretation, the model initially encodes every image into a visual attribute index by withdrawing the relative attribute plots with CNN, and later exploits RNN to decode the visual attributes to a pattern of words. The simple explanation for this has been represented in the fig. 10. Majority of DL segmentation approaches operate on encoder-decoder archetype that can be divided into 2 sections: encoder-decoder framework for universal segmentation, and for therapeutic segmentation.

2.1.6.4. Encoder-decoder approaches for general segmentation. Hyewon-woo et al. [210] presented an innovative idea on semantic segmentation founded on de-convolution. The framework comprises 2 components, an encoder and a de-convolutional structure. The de-convolution arrangement is achieved by using de-convolution and un-pooling layers, that discovers element-wise hierarchy labels and anticipates segmentation masks. This approach attained probable results with the outstanding accuracy of 72.5 % amid the procedures trained without any extraneous content at that period. A supplementary advantageous approach presented the similar technique as SegNet [182]. In another study a Bayesian form of SegNet was introduced for simulating the ambiguity essential for the convolutional encoder-decoder network [211]. High-resolution Network (HRNet) is important framework belonging to this classification [212]. HRNet sustains high-resolution depictions over the encoding procedure by associating high to low resolution convolution streams simultaneously, and continually switching the data around resolutions. A couple of other approaches that endorse inverted convolutions for image segmentation are Stacked De-convolutional Network (SDN) [213], LinkNet [214], W-Net [215].

2.1.6.5. Encoder-decoder models for therapeutic image segmentation. There exist number of prototypes primarily implemented for therapeutic image segmentation that are derived from FCNs and encoder-decoder architectures [216,217]. The U-Net is predominant architecture employed in segmenting therapeutic data and is composed of 2 portions: a decimation to take context, and a proportional extending track that

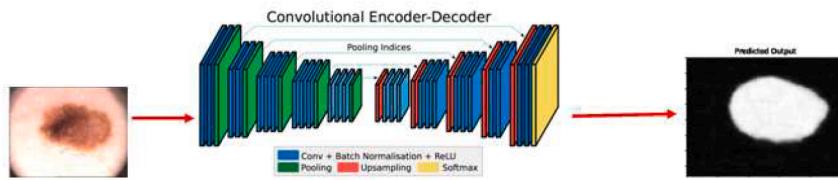


Fig. 12. The architecture of encoder-decoder model for therapeutic images [223]

authorizes explicit localization. The decimation portion has a FCN-like framework that withdraws attributes with 3×3 convolutions. The latter portion employs up-convolution, mitigating the amount of working model while expanding their magnitude. Miscellaneous expansions of U-Net have been evolved for diverse forms of representations as Ozgun et.al [218] presented this framework for 3D data. Zongwei et.al. [219] proposed an embedded form of this framework. It has also been exercised on numerous matters like a road withdrawal computation predicted on UNet [220]. The architecture of U-Net approach is illustrated in fig. 11. Similarly one more prominent FCN based framework is V-Net for 3D therapeutic segmentation of illustrations [217,222]. Here, for learning model, a novel decision criterion as stated by Dice coefficient (DI), facilitating the architecture to handle circumstances where there is a persistent asymmetry among the counts of voxels in the forefront and background. The framework was practiced throughout on MRI degree of gland, and assimilates to foresee segmentation for the entire degree instantaneously. The architecture of encoder-decoder based segmentation approach is illustrated in fig. 12.

2.1.6.6. Pyramid network based models (PN). This evaluation is quiet an old theory in image refining that has been implemented in numerous NN frameworks. Amongst the most notable networks of the arrangement is the Feature PN (FPN) that was initially designed for recognizing objects but was later on utilized in segmentation process [224]. The pyramidal distribution of deep CNNs is operated to formulate attribute stacks with negligible additional expense. For allying low- and high-resolution attributes, the FPN is formed from a bottom-up, a top-down and adjacent orientations. The concurrent model system is then transformed by a 3×3 convolution to generate the production of every level. At last, every level of the top-down lane creates a speculation to determine an entity. The implementation of an arrangement called Pyramid Scene Parsing Network (PSPN) to ascertain the comprehensive frame delineation of a scene is discussed in [225]. Various samples are withdrawn from the loaded image employing a ResNet as an attribute withdrawal, by a distended framework. The design types are later sustained to a pyramid pooling constituent to determine sequence of diverse hierarchies. A multi-resolution transformation configuration derived from a Laplacian pyramid that utilizes shortcut networking from higher definition design types and augmentative convergence to efficiently fine-tune fragment confines reassembled from lower-resolution layouts is presented in [226]. It was revealed that, even if the recognizable dimensional pixel count of working model is poor, the high-definition attribute depiction accommodates remarkable sub-pixel localization data. Some other representations that utilize similar methodology for segmentation is Dynamic Multi-scale Filters Network DM-Net [227]. Context data is recognized as vital for scene labeling which considerably enhance the interpretation leading to Context Contrasted Network (CCN) [228]. Another methodology that produces multi-scale contextual depictions with innumerable highly developed ACMs is Adaptive PCN (APC) [229]. Multi-scale context intertwining (MSCI) [230] and salient entity segmentation [231] are amongst another innovations.

2.1.6.7. Regional CNN model. This procedure has demonstrated its effectiveness in identifying the entities as an application. Not only R-

CNN, but its extension (Faster R-CNN approach) is evolved for identifying the entities to bring forward bounding box entrants [232]. It withdraws targeted regions and corresponding pool layer calculates attributes out of these initiations in exchange for deducing the bounding box coordinates and the family of the entity. A few types of this procedure are extensively employed to tackle the segmentation issues like the task of producing entity identification and segmentation concurrently. Kaiming et al. [233] recommended a Mask R-CNN network which effectively identifies entities in a representation and at the same time produces a superior segmentation mask for every specific case. It is basically a Faster RCNN possessing 3 output sections—constituting the bounding box points of reference, the corresponding classifications, and the binary mask to partition the entity respectively. Another approach established on this architecture is the Path Aggregation Network (PANet) [234]. The attribute withdrawal of the system employs FPN framework with a novel dilated bottom-up trail enhancing the distribution of low-layer attributes. The result is combined to the analogous phase design type of the top-down trail employing an edgewise association and sustains the following step. The preceding subdivision utilizes a FC layer to produce the projections of the bounding box points of reference and the corresponding entity type. The last one refines the targeted region with FCN to predict the entity mask. Jifeng et al. [235] presented an architecture for instance segmentation, which encompasses 3 grids, discretely separating instances, evaluating masks, and classifying entities. Such arrangement configures a cascaded organization, and is fabricated to allocate their convolutional attributes. Ronghang et al. [236] recommended a novel relatively-supervised training prototype, besides an unprecedented weight correction factor, which entitles training instance segmentation prototypes on a significant variety of groups, each comprising box interpretations. Liang et al. [237] proposed instance segmentation architecture by revising entity recognition with semantic attributes derived from Faster R-CNN. The prototype generates 3 results: box identification, semantic segmentation, and orientation projection. Tensormask is one more innovative approach that is formulated on concentrated sliding window object segmentation serving as a projection undertaking over 4D tensors and withdraws a conventional configuration [234]. Numerous approaches relying on instance segmentation have been produced like R-FCN [238], DeepMask [239], PolarMask [240], boundary-aware instance segmentation [241], CenterMask [242], Deep WT [243], real-time instance segmentation [244], and Semantic Instance segmentation via Deep Metric Learning [245].

2.1.6.8. Atrous convolutional models. Atrous convolution instigates specification to convolutional layers i.e., the dilation rate. The procedure has been approved in the domain of real-time segmentation and numerous contemporary works surveyed the deployment of this methodology. Out of these, the prominent one is the DeepLab [197]. It has developed gradually to v1, v2, v3, and, currently, v3+. Atrous spatial pyramid pooling (ASPP) is comprehended in v2 where the attributes are accumulated after being refined by means of atrous convolutional layers of varying dimensions. Context assemblage [246], dense upsampling convolution and hybrid dilated convolution (DUC-HDC) [247], densely connected ASPP (DenseASPP) [248], and the ENet [249] are among



Fig. 13. Original image, original label and deeplab V3 segmented images (<http://awesomopensource.com/project/lattice-ai/DeepLabV3-Plus>)

other innovations. Other than above mentioned approaches DeepLabv1 and DeepLabv2 [197] are also effective innovations in image segmentation. DeepLabv2 has 3 important characteristics: the employment of dilated convolution for examining the diminishing resolution in the framework; ASPP that analyzes a designated convolutional layer with screens at varied duty cycles thereby apprehending entities and image context at varied ranges to partition entities at varied scales; and rectified positioning of entity margins by fusing mechanisms from deep CNNs and stochastic graphical frameworks. A Deeplabv3+ that employs encoder-decoder organization, incorporating atrous dissociable convolution which is formed from a depth and point-based convolution utilizing deeplabv3 as encoder is discussed in [250]. The deeplab V3 based segmentation outcomes are presented in fig. 13.

2.1.6.9. Recurrent neural network based models. RNNs are functional in approximating the transient relativity amongst pixels to enhance the evaluation of the distribution layout. The pixels are interconnected and refined serially to represent multidisciplinary forms and to develop semantic segmentation by applying RNN (fig. 14). A novel architecture formulated on RNN and ReNet is ReSeg that is employed for semantic segmentation [251,252]. Every ReNet layer encompasses 4 RNNs that glides the image in each direction encoding stimulations and supplying appropriate data.

LSTM networks are other kinds of RNN which can assimilate order dependence. The resultant of the former stage is utilized as intake in the ongoing stage in RNN. The LSTM might keep data for extended period by its definition and is employed for time-series data refining, and categorization. The utilization of LSTM approach for image segmentation and categorization has been discussed in [253] examining 2D LSTM

strategies for images of the complex dimensional relativities of labels. An innovative approach for semantic segmentation architecture in accordance with the Graph LSTM methodology is presented in [254]. Another innovation in the domain of RNN methodology is the evolution of Data Associated RNN (DA-RNNs) projected for combined 3D scene plotting and semantic marking [255]. A fusion of CNN and LSTM based approach is proposed relying on semantic segmentation formula for encoding the image and its natural language specification [256]. A series of recurring NN units comprises all RNN and this unit in typical RNNs will have a basic framework such as an isolated tanh layer. The resultant of the existing time step turns the input for the subsequent time step, which is considered as Recurrent. At respective constituent of the series, the prototype analyzes not only the existing input, but also what it discovers about the preceding ones. An isolated layer prevails in the recurring unit of a conventional RNN.

2.1.6.10. Attention-based models. This implementation has been investigated in CV, and the applications which employ these procedures are found in semantic segmentation. An invention suggested on this procedure that ascertains to import multi-scale attributes at every pixel point is depicted in [258]. A procedure called Reverse Attention (RAN) is developed that prepares the model to attain the reverse notion

[259]. The network is partitioned into 3 divisions which carry out the straight and reverse cognitive procedure at once. A PANet approach implemented for semantic segmentation that employs the effect of comprehensive contextual data is represented in [260]. In this work a fusion of attention technique and spatial pyramids is employed for withdrawing accurate characteristics for pixel marking, rather than complex dilated networks. In another approach, a dual attention network for segmentation is deployed [261]. In this method, 2 kinds of attention units are adjoined that represents the semantic coordination in dimensional and medium proportions. There are numerous related techniques relying on this procedure like OCNet [262], Expectation-Maximization ANet (EMANet) [263], Criss-Cross ANet (CCNet) [264], end-to-end segmentation with RAN [265], pointwise spatial ANet (PSANet) [266], and a discriminative feature (DFN) [267].

2.1.6.11. Generative adversarial network (GAN) approach. This procedure has been implemented extensively in CV, and has been assimilated for image segmentation as well [268,269]. A GAN formulated technique trained on convolutional semantic segmentation architecture distinguishing ground-truth segmentation maps from those produced by the segmentation arrangement is presented in [270]. In another study a semi-weakly directed segmentation employing GAN is presented [271]. A similar approach was presented in [272] utilizing FCN to distinguish the predicted maps and the ground truth segmentation arrangement. A

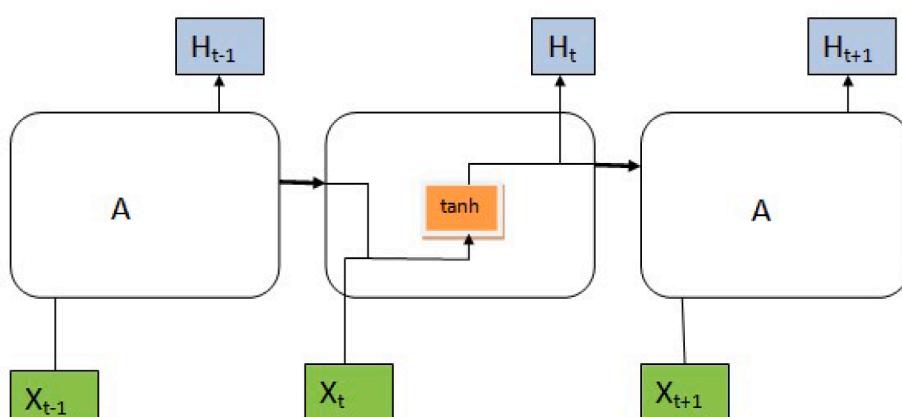


Fig. 14. Illustration of recurrent NN process [257]

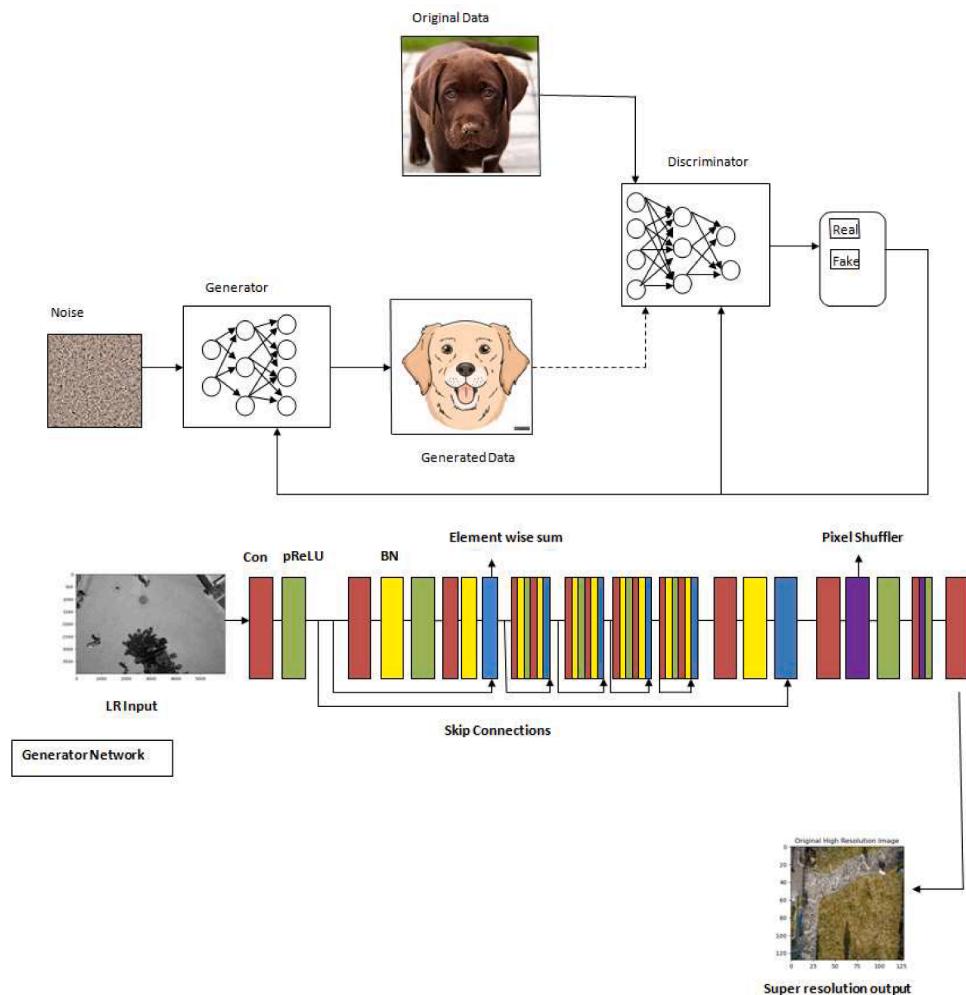


Fig. 15. (a). Illustration of GAN based segmentation process [274] (b). Illustration of Generator network for GAN based approach [275] (c). Illustration of Discriminator network for GAN based approach [275]

GAN with multi-scale loss for therapeutic segmentation of data was developed using a FCN and projected an innovative adversarial evaluator system [273]. Fig. 15(a) illustrates basic GAN architecture, 15(b) illustrates generator network's working and 15(c) illustrates discriminator network's working.

2.1.6.12. Active contour approach. The research of association among FCN and ACM has lately drawn attention of the analysts. A supervised loss layer is generated comprising details of the predicted masks to resolve the complications of orifice segmentation in MRI [276,277]. An application of the same has been depicted where the level-set ACM is executed as RNN [278]. Additionally, Deep Active Contour (DAC) is one more innovative technique that falls under this category [279]. A further extension in this domain includes Deep Active Lesion Segmentation (DALs) strategy for therapeutic images [280]. A further innovation includes Deep Structured AC (DSAC) that fuses ACM and FCN to assemble instance segmentation [281]. Dominic et al. recommended an alternate approach for the similar practice but with distinct ACM design called Deep Active Ray Network (DarNet) [282]. An accurate sequential BPN trainable, fully-integrated fusion of FCN and ACM known as Dubbed DCAC (DCAC) has been presented by Ali et al. [283].

2.1.6.13. Other Models. Some other DL methodologies for segmentation are: Context Encoding Network (EncNet) which utilizes a fundamental extraction of characteristics [284]. Another technique includes a fine-tuning system i.e, RefineNet which precisely utilizes all the data

accessible across the down-sampling procedure to validate high-resolution projection exploiting extended residual associations [285]. A different approach that instigated a self-regulating seed initiation procedure accompanied by augment learning to resolve the issue of interactive segmentation is SeedNet [286]. A yet another technique that comprehends and estimates the entity domains based on the ground truth is the "Object-Contextual Representations" (OCR) [212]. Another representation is BoxSup that is a bounding box interpretation as a substitute or additional source of observation to train CNN for semantic segmentation [287]. Similarly further technique that is employed for negotiating the segmentation and localization challenges is Graph CNN (GCN) [288]. Another innovation possessing superficial framework of residual networks that considerably excels substantial deeper models is Wide ResNet [289]. One more strategy called ExFuse has been developed to overcome the differences among low and high-level characteristics and to considerably enhance the segmentation quality [290]. A distinct technique formulated merely on feed forward design for semantic segmentation is 'Feed Forward Net' [291]. An alternative mechanism that employs saliency as antecedent for entity by means of the estimation of geodesic dimension is saliency-aware model [292]. One more methodology that comprises 2 supplementary learning challenges: a) estimating label maps from images, b) modifying the images utilizing these label maps which are collectively unraveled is Dual Image Segmentation (DIS) [293].

Another methodology which evaluates the contour configuration of a scene image via a CNN that assimilates evidence from contextual entities

in the image is FoveaNet [294]. A novel technique formulated on DenseNet-based framework that characterizes prominent tuning capability and an un-sampling data route is Ladder DenseNet [295]. In another approach, Bilateral Segmentation Network (BiSeNet) is presented that comprises 2 segments: Spatial and Context Path (CP) implemented to tackle the loss of dimensional data and contraction of receptive field [296]. One more model that assimilates to re-weight the confined attributes by means of the supervision from pixel-wise semantic estimation is referred to as Semantic Prediction Guidance (SPGNet) [297]. A novel 2 cascaded CNN approach for semantic segmentation which distinctly wires contour detail as an autonomous refining division is another innovation and is called Gated-SCNN [298]. Another technique introduced to seize the pixel based contexts through a synthesis of comprehensive and local context as determined by diverse pixel requirements is called an Adaptive Context Network (ACNet) [299]. A methodology that assembles a semantic neuron plot by specifically integrating the semantic hypothesis spectrum into network assembly is known as Dynamic Structured Semantic Propagation Network (DSSPN) [300]. A distinct methodology that exercises rationale around a cluster of metamorphic nodules whose outcomes depict individual characteristics of every semantic in an existing knowledge plot is known as Symbolic Graph Reasoning (SGR) [301]. For managing the long-tail dispersion of entities in scenes and the dimensional configuration connections of scenes, entities, and segments of entities, a technique called CascadeNet [132] is deployed [302]. Another approach implemented to modify the magnitudes of receptive fields based on the magnitudes of entities is Scale-Adaptive Convolutions (SAC) [303]. Another model that is exploited as the acknowledgement of various graphic principles as feasible from a specified image is defined as Unified Perceptual Parsing (UperNet) [304]. A yet another segmentation procedure by re-training and self-training is developed in [305]. For authorizing system frameworks and to prevent the discrepancy among practical and substitute context, Densely Connected Neural Architecture Search(DCNAS) is deployed [306]. A hierarchical methodology where rather than learning all attention masks for all rigid form of levels, learning a proportionate attention mask among contiguous ranges is done [307]. Panoptic segmentation (PS) [308] is another segmentation approach with increasingly widespread use and its other extensions include PS Feature PAN [309], ANet PS [310], Seamless Scene Segmentation [311], PS Deeplab [312], UPSNet [313], EfficientPS [314].

3. Datasets for image segmentation

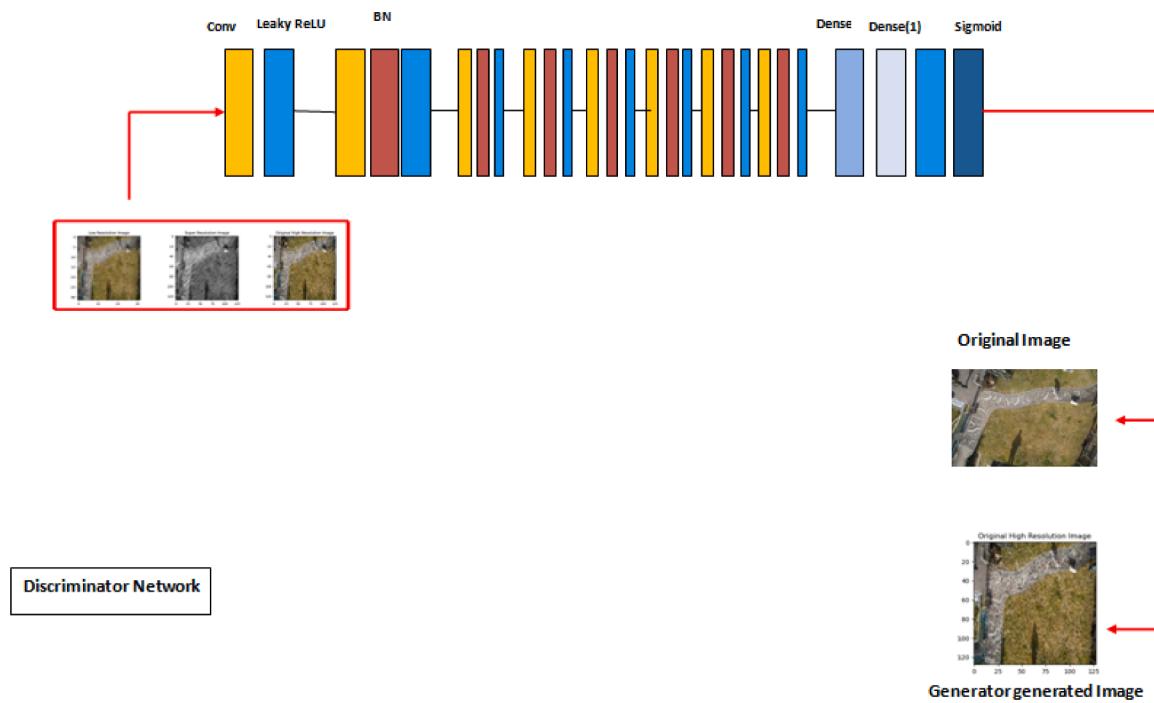
In this section an overview of extensively utilized image segmentation databases for 2 D images is presented.

3.1. Non-therapeutic image datasets

The maximum number of segmentation of data analysis has settled on 2D pictures; so, numerous 2D segmentation databases are present such as PASCAL Visual Object Classes (VOC) [315]. This is amongst the widely used datasets in CV, with illustrated images accessible for 5 problems—categorization, segmentation, identification, diagnosis, and person mapping. PASCAL Context [316] is a progression of the PASCAL VOC 2010 dataset. It comprises labels exceeding 400 that are partitioned into 3 types. The additional wide ranging dataset for detecting objects, segmenting the pictures, and picture labelling is Microsoft Common Objects in Context (MS COCO) [317]. COCO contains pictures of composite common views, encompassing ordinary entities. The dataset

comprises pictures of 91 commodities group, and a sum of 2.5 million designated samples in 328000 pictures. Similarly another wide ranging dataset having an emphasis on semantic apprehension of urban street view is Cityscapes [318]. It accommodates a distinctive arrangement of virtual video arrangements produced in road views. It incorporates 30 categories; partitioned into 8 types—level surface, persons, automobiles, buildings, entities, nature, airspace, and barren. One more dataset utilized for conventional training and estimation for scene parsing computation is ADE20K /MIT Scene Parsing [302]. The dataset comprises more than 20,000 scene fundamental pictures illustrated with entities and its parts and 150 semantic classifications. The instance is categorized into 20,000 pictures for training, 2000 for validation, and others for testing. Another dataset that incorporates 2688 illustrated pictures from a division of the LabelMe dataset is SiftFlow [319]. The pictures having 256×256 -pixel values are formulated on 8 diverse outside views, particularly road, hills, grassland, coasts, and constructions. All these pictures are associated with 1 of 33 semantic categories. One more dataset that comprises outside view pictures from prevailing datasets is Stanford background [320]. The dataset incorporates 715 pictures with 1 foreground entity in minimum.

A dataset of human segmented Natural pictures incorporating 12,000 pictures is Berkeley Segmentation Dataset (BSD) [321]. The objective is to issue an experimental foundation for analysis on image segmentation and edge identification. A fraction of the segmentations was procured from displaying the content a RGB picture and another fraction from displaying a gray picture. Youtube-Objects [322] accommodates videos extracted from YouTube, that incorporate entities from 10 PASCAL VOC categories. KITTI [323] is another dataset for mobile automations and self-contained vehicle. It incorporates time unit videos of vehicle scenes, produced by a range of detectors such as cameras. The benchmark holds no special base for semantic segmentation, but analysts have artificially elucidated portions of benchmark for exploration objective [324]. The DUTS image dataset is contemporarily the substantial saliency recognition standard with the precise training/test consideration regimen comprising 10,553 training pictures and 5,019 test pictures [325]. Cornell iCoseg Dataset: It is an extensive open source co-segmentation dataset of 38 groups i.e., 643 pictures [326]. NYU Depth Dataset V1: The dataset is encompassed of video progressions from a collection of indoor locales as produced by the RGB and Depth cameras [327]. Matting Human Datasets: This is the sizeable vignette matting dataset, encompassing 34,427 pictures and analogous matting outcomes [328]. Pedestrian Parsing Dataset: This is a Deep Decompositional Network (DDN) based dataset for interpreting wayfarer pictures into semantic zones. The methodology plots low-level ocular contemplations to the label maps of anatomical parts with DDN that is capable of precisely evaluating convoluted pose disparities with fine robustness to constrictions and background scatters. DDN collectively evaluates constricted areas and partitions body parts by accumulating 3 kinds of hidden layers: constricted evaluation, termination, and disintegration layers [329]. Instance Segmentation in Aerial Images (iSAID): iSAID is foremost standard database for instance segmentation in aerial pictures. The exceedingly large and densely annotated benchmark comprise 655,451 entity instances for 15 classes [330]. Semantic Drone Dataset: It emphasizes semantic assimilation of urban outlook for augmenting the security of self-governing drone trajectory and landing mechanisms. The illustration represents residences exceeding 20 from overhead perspective [331]. More datasets for image segmentation are Semantic Boundaries Dataset (SBD) [332], PASCAL Part [333], SYNTHIA [334], and Adobe's Portrait Segmentation [335].



3.2. Therapeutic image datasets

ISIC: The International Skin Imaging Collaboration (ISIC) comprises 2357 pictures of skin cancer. The pictures are arranged so as to categorize the skin cancer into 8 types with same number of images except for melanoma. **Chest X-Ray (Pneumonia):** The dataset is arranged into 3 folders namely train, test and validation. Each folder comprises sub folder for each image classification as normal or pneumonia. **PH2 dataset:** This dataset comprises a total of 200 dermoscopic pictures, having 80 common nevi, 80 atypical nevi, and 40 melanomas. **Liver Tumour Segmentation (LiTS) dataset:** The training data set comprises 130 CT scans and the test data set 70 CT scans [336]. **IOSTAR Retinal Vessel Segmentation Dataset:** It comprises 30 pictures having a pixel count of 1024×1024 . Every vessel is interpreted by a class of specialists engaged in the domain of retinal image evaluation. Furthermore the benchmark comprises interpretations for the blind spot and the artery/vein ratio. [337]

3.3. 4 Evaluation metrics

In this section, an outline of recommended criterion operated in estimating the execution of segmentation images, and the analysis of the segmentation techniques on different databases is presented. When completing a binary classification task, data examples are usually projected to be either positive or negative. This means that a positive label indicates the occurrence of a disease, abnormalities, or other variations, whereas a negative label indicates that the data instance is similar to the baseline. As a result, there are 4 probable distinctions for each estimated binary tag: false positive (FP) denotes an instance that was incorrectly anticipated to be positive, true negative (TN) denotes an accurately anticipated negative result, and true positive (TP) denotes an accurately anticipated favorable result.

Accuracy, sensitivity, specificity, and precision are among the most often used metrics of assessment for binary classification. They indicate, respectively, the number of cases correctly identified in the set of all instances, the truly positive cases, the truly negative cases, and those that are categorized as positive. Recall is a popular term used to describe sensitivity. A number of additional assessment measures, such as accuracy, are likewise dependent on every value in the confusion matrix:

Sen. + Spe. – 1 is the definition of Youden's index, which assigns equal weight to the accuracy of both positive and negative examples, despite their respective quantities. When comparing the performance of the binary classifier to the randomized accuracy, the kappa score is utilized. Although it can be used to gauge the extent of agreement amongst projected and actual categories, it was first developed as a psychological measure of the percentage of similarity among 2 observers. Additionally, the definition of the Matthews' correlation coefficient (MCC) evaluates the relationship among the instances' actual and expected values. The mathematical description of Pearson's correlation coefficient correlates with this one regarding MCC. The framework must apply a threshold to the test set hypotheses in order to turn them into binary labels, in order to calculate the values of the aforementioned metrics. Youden's index, the cut-point that yields the maximum accuracy or the standard deviation of 0.5 for the training set's projection is frequently used as the threshold value. It is imperative to select the threshold solely according to the training set's estimations, as selecting the threshold that optimizes the test set's predictions for accuracy leads to absurdly favorable outcomes. On the other hand, the receiver operating characteristic (ROC) curve can be taken into consideration if the numerical projections prior to their translation into binary are known. Calculating sensitivity against the false positive rate—which is equal to 1 minus specificity—at every threshold value that can be chosen yields this result. The ROC curve is a gradually growing curve inside the unit square that is connected to the points $(0, 0)$ and $(1, 1)$. The more close the ROC curve is to $(0, 1)$, the more accurate predicted values are. Another potential assessment metric with values in the range of 0 to 1 is the area under the ROC curve (AUC). However, in contrast to the other metrics, the value of AUC is independent of the threshold selection. **Multi-class classification:** A $m \times m$ confusion matrix, as can be used to display the classifier's output if the aim is to divide n occurrences among distinct classes. The number of cases from the i th categorized to the j th class is represented by its element n_{ij} at the coordinates of the i th and the j th column for $i, j = 1, \dots, l$. The same metrics that we described for binary classification are used in the evaluation of this matrix.

First off, each of the metric variables may be obtained in two easy ways, with the exception of the AUC that was presented in the previous section. For every k class, a distinct 2×2 confusion matrix must be made. Additionally, Cohen's κ and MCC have their own definitions

created specifically for multi-class classification. Class labelling that are not exclusive are used in multi-label classification, which is an extended version of multi-class classification. The goal is to identify all the class labels that apply out of $k \geq 2$ available labels, as opposed to splitting the data samples among many classes. The model outputs a binary vector \mathbf{y} (i) for each n instances, where $i = 1, \dots, n$. For every $j = 1, \dots, k$, the jth element is 1 if the jth label is present and 0 otherwise. One potential evaluation statistic is the Hamming loss.

3.4. Performance criterion for segmentation models

A model should be assessed in several estimations, including quantitative accuracy, hypothesizing and memory specifications. But majority of investigation study so far, revolve around the parameters for estimating the model efficiency. Therefore, we present the approving parameters for evaluating the reliability of segmentation procedures (Table 1).

3.4.1. Pixel accuracy (PA)

It directly detects the ratio of pixels accurately identified, to the total number of pixels. For $M + 1$ categories PA is defined as Eq. (20):

$$PA = \frac{\sum_{i=0}^M P_{ii}}{\sum_{i=0}^M \sum_{j=0}^M P_{ij}} \quad (20)$$

P_{ij} represents number of predicted pixels of category i belonging to class j.

3.4.2. Mean pixel accuracy (MPA)

It is the enhanced form of PA, where the ratio of accurate pixels is calculated in a per-class mode and later averaged over the aggregated categories. MPA can be defined as in equation 21:

$$MPA = \frac{1}{M+1} \sum_{i=0}^M \frac{P_{ii}}{\sum_{j=0}^M P_{ij}} \quad (21)$$

3.4.3. Intersection over union (IoU)

The predominantly experimented parameters in semantic segmentation is IoU. It is characterized as the region of convergence among the predicted segmentation plan and the ground truth, to the area of union among the predicted segmentation and the ground truth (equation 22).

$$IoU = J(X, Y) = \frac{|X \cap Y|}{|X \cup Y|} \quad (22)$$

X, Y represents ground truth.

3.4.4. Mean-IoU

This is distinct extensively utilized measure, that is characterized as the standard IoU around each category. It is employed in characterizing the evaluation of contemporary segmentation techniques.

3.4.5. Mean square error (MSE)

MSE measures the average of squared divergence among the defined score and the estimated score by the model. The error among 2 images $f(m,n)$ and $f'(m,n)$ can sometimes be negative and to prevent this, MSE is calculated (equation 23).

$$MSE = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad (23)$$

x_i refers to ground truth. \hat{x}_i refers to estimated score n = amount of data.

3.4.6. Mean absolute error (MAE)

It refers to the average of divergence among ground truth and estimated scores (equation 24).

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i| \quad (24)$$

x_i refers to ground truth. \hat{x}_i refers to estimated score n=number of data

3.4.7. Universal quality index (UQI)

The UQI is given by the equation 25.

$$UQI = \frac{4\sigma_{mn}\bar{m}\bar{n}}{(\sigma_m^2\sigma_n^2)((\bar{m})^2 + (\bar{n})^2)} \quad (25)$$

\bar{m} and \bar{n} , represents mean values of input and noisy images respectively. σ_m^2 and σ_n^2 , represents the variance σ_{mn} , represents covariance

3.4.8. Root mean square error (RMSE)

It refers to the square root of average of squared divergence among the expected and estimated score by the model (equation 26).

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i| \quad (26)$$

3.4.9. R^2 Coefficient

The calculation of R^2 is done using different metrics and it measures what percentage of dissimilarity in target (X) is interpreted by the dissimilarity in regression line (Y). It is measured by sum of square error (SE) given by the equation (27):

$$SE(\bar{X}) = \sum_{i=1}^n (x_i - \bar{x})^2$$

Percentage of dissimilarity reported by regression line is given by the equation:

$$1 - \frac{SE(\text{line})}{SE(\bar{X})}$$

Ultimately, the formula for R^2 is computed using the equation:

$$\text{coeff}(R^2) = 1 - \frac{SE(\text{line})}{SE(\bar{X})} \quad (27)$$

3.4.10. Adjusted R^2

In some methodologies, the analysts are deceived by supposing the model to be improving when the value increases while no learning is taking place (over fitting). For correcting this issue, R^2 is modified with some independent variables. Its value will always be less than R^2 and is given by the equation (28):

$$R_A^2 = 1 - \left[\left(\frac{m-1}{m-n-1} \right) \cdot (1 - R^2) \right] \quad (28)$$

m = number of observations. n=number of independent variables

$$R_A^2 = \text{Adjusted } R^2$$

3.4.11. Gradient similarity metric (GSM)

It measures local quality map generating single quality value by utilizing standard deviation. Of all the image quality magnitudes, GSM is fastest in computation and is defined by the equation (29):

$$GSM = \frac{2x_a x_b + m}{x_a^2 + x_b^2 + m} \quad (29)$$

x_a and x_b , represents the GM of input and noisy images respectively

By utilizing the standard deviation as pooling method, it is referred to as GMS Deviation or GMSD and is represented by the equation (30):

$$GMSD = \sqrt{\frac{1}{xy} \sum_{k=1}^x \sum_{l=1}^y (GMS(k,l) - GMSM)^2} \quad (30)$$

$$GMSD \propto \frac{1}{Q}$$

Q= Image quality

3.4.12. Precision (Pre)

It is also known as Positive Predictive Value (PPV) that calculates the number of accurate predictions of all the positive predictions (equation 31).

$$Pre, PPV = \frac{tp}{tp + fp} \quad (31)$$

TP= True Positive. TN= True Negative. FP= False Positive

FN= False Negative

3.4.13. Precision-recall curve

It refers to a curve associating the PPV and TPR to a single solution.

3.4.14. Average precision

It is also known as PR AUC and can be defined as the aggregate of precision values measured for every recall value (equation 32).

$$AP = \sum_{n=0}^{k-1} Recall(n) - Recall(n+1)*Pre(n) \quad (32)$$

3.4.15. Recall

It is also known as True Positive Rate (TPR) that measures the number of positive observations classified of all positive observations (equation 33).

$$TPR, Recall = \frac{TN}{TP + FN} \quad (33)$$

3.4.16. F1 score

It is the harmonic mean of precision and recall given by equation (34).

$$F1 = \frac{2 * Pre * Recall}{Pre + Recall} \quad (34)$$

3.4.17. F2 score

A combination of precision and recall that intensifies the recall by 2 times given by equation (35).

$$F2 = \frac{((1 + 2^2)^* Pre * Recall)}{(2^2 * Pre + Recall)} \quad (35)$$

3.4.18. False positive rate (FPR)

FPR can be characterized by a fraction of fake alerts that is going to augment on the basis of model predictions. It is also called type 1 error given by equation (36).

$$FPR = \frac{fp}{fp + tn} \quad (36)$$

3.4.19. False negative rate (FNR)

FNR can be characterized by a fraction of neglected false transactions that the model introduces. It is also known as type 2 error given by equation (37).

$$FNR = \frac{fn}{tp + fn} \quad (37)$$

3.4.20. Negative predictive value (NPV)

It calculates the number of accurate predictions of all the negative predictions given by equation (38).

$$NPV = \frac{tn}{tn + fn} \quad (38)$$

3.4.21. False discovery rate (FDR)

It calculates the number of inaccurate predictions of all the positive predictions given by equation (39).

$$FDR = \frac{fp}{fp + tp} \quad (39)$$

3.4.22. Accuracy

It calculates if positive and negative predictions are accurately classified given by equation (40).

$$ACC = \frac{tp + tn}{tp + fp + tn + fn} \quad (40)$$

3.4.23. Dice coefficient

This is one more important metric for image segmentation that is employed mostly in therapeutic image evaluation. It is specified as double the overlap region of anticipated and special base layouts, by overall pixels in both representations given by equation (41).

$$Dice = 2 \frac{|X \cap Y|}{|X \cup Y|}, \quad (41)$$

$$Dice = \frac{2TP}{2TP + FP + FN}$$

3.4.24. Sensitivity

It is the estimation of a positive test outcome, depending on the particular truly being positive. It is also known as True Positive Rate (TPR) given by equation (42).

$$SE = \frac{TP}{TP + FN} \quad (42)$$

3.4.25. Specificity

It is the estimation of a negative test outcome, depending on the particular truly being negative. It means of all the negative outcomes; how many have been classified as negative. It is also known as true negative rate (TNR) and is given by equation (43).

$$SP = \frac{TN}{TN + FP} \quad (43)$$

3.4.26. F-Beta score

It is a combination of precision and recall and the higher value of it indicates that the model performs better given by equation (44).

$$F_B = (1 + \beta)^2 \frac{Pre * Recall}{\beta^2 * Pre + Recall} \quad (44)$$

3.4.27. Cohen's kappa score

For calculating it, 2 implementations are utilized namely ‘observed agreement’ (P_a) and ‘expected agreement’ (P_b). P_a characterizes how the estimations of a classifier agree to the ground truth. P_b characterizes how the estimations of arbitrary classifier agree to the ground truth (equation 45).

$$\kappa = \frac{P_a - P_b}{1 - P_b} \quad (45)$$

$$\kappa = \frac{2(TP \times TN - FN \times FP)}{(TP + FP) \times (FP + TN) + (TP + FN) \times (FN + TN)}$$

Table 1
Summary of all the evaluation metrics for image segmentation

Performance Metrics	Formula	Description	Application
Recall	$TPR, Recall = \frac{TN}{TP + FN}$	It measures the number of positive observations classified of all positive observations.	<ul style="list-style-type: none"> • It is used in case of imbalanced data.
False Positive Rate	$FPR = \frac{fp}{fp + tn}$	A fraction of fake alerts that is going to augment on the basis of model predictions	<ul style="list-style-type: none"> • It is usually used as an additional metric and not solely. • By augmenting the threshold values, fewer alerts are obtained. • It is usually used as an additional metric and not solely. • It is usually used as an additional metric with precision. • It is a go-to metric that detects all false executions at the cost of false alerts. • It is used to ensure that the decision made is accurate.
False Negative Rate	$FNR = \frac{fn}{tp + fn}$	A fraction of neglected false transactions that the model introduces. It is also known as type 2 error.	
Sensitivity	$SE = \frac{TP}{TP + FN}$	It is the estimation of a positive test outcome, depending on the particular truly being positive.	
Specificity, TNR	$SP = \frac{TN}{TN + FP}$	It is the estimation of a negative test outcome, depending on the particular truly being negative. It means of all the negative outcomes, how many have been classified as negative.	
Negative Predictive value	$NPV = \frac{tn}{tn + fn}$	It calculates the number of accurate predictions of all the negative predictions.	<ul style="list-style-type: none"> • It is utilized when higher precision is required on negative estimations.
False Discovery Rate	$FDR = \frac{fp}{fp + tp}$	It calculates the number of inaccurate predictions of all the positive predictions.	<ul style="list-style-type: none"> • Again, it is used as an additional metric for example recall. • It is utilized in case of considering all the positive estimations to be worth noticing. • It is utilized in case realizing the false alerts is costly.
Precision	$Pre, PPV = \frac{tp}{tp + fp}$	It is also known as Positive Predictive Value (PPV) that calculates the number of accurate predictions of all the positive predictions.	
F Beta	$F_B = (1 + \beta)^2 \frac{Pre * Recall}{\beta^2 * Pre + Recall}$	It is a combination of precision and recall and the higher value of it indicates that the model performs better.	<ul style="list-style-type: none"> • The higher value of β is to be chosen in case recall value is considered important over the precision. • Whereas if precision is to be considered crucial, $0 < \beta < 1$ is followed. • It is utilized in all binary classifications.
F 1 score	$F_1 = \frac{2Pre \cdot Recall}{Pre + Recall}$	It is the harmonic mean of precision and recall.	
F 2 score	$F_2 = \frac{2Pre \cdot Recall}{Pre + Recall}$	It is the harmonic mean of precision and recall.	
Cohen Kappa	$\kappa = \frac{2(TP \times TN - FN \times FP)}{(TP + FP) \times (FP + TN) + (TP + FN) \times (FN + TN)}$	It determines the superiority of the model over the arbitrary classifier which estimates the performance on the basis of class frequency.	<ul style="list-style-type: none"> • It is used when recalling positive observations is crucial than being precise about it. • It operates well in case of imbalanced data.
Matthews Correlation Coefficient	$MCC = \frac{(TP \times TN - FP \times FN)}{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}$	The association among estimated class and the ground truth is referred to as MCC.	
ROC		It refers to a 2 D graph that predicts the tradeoff among TPR and FPR.	
ROC-AUC		ROC-AUC determines a number that represents how accurate the curve is.	<ul style="list-style-type: none"> • It is used in case of ranking predictions. • It cannot be used in case of imbalanced data. • It is used to come to a conclusion in case of precision/recall difficulty. • It is used when data is imbalanced.
Precision-Recall Curve		It refers to a curve associating the PPV and TPR to a single solution.	
Average Precision	$AP = \sum_{n=0}^{k-1} Recall(n) - Recall(n+1) * Pre(n)$	It is defined as the aggregate of precision values measured for every recall value.	
Mean Squared Error	$MSE = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2$	MSE measures the average of squared divergence among the defined score and the estimated score by the model	<ul style="list-style-type: none"> • It is optimized better because of being differentiable. • It reports even the smallest erroneous values by squaring them. • It is used where error evaluation doesn't need second thought. • It is used to maintain differentiable attribute of MSE • It is used to administer penalization of miscalculations presented by MSE by square rooting them. • It determines how well the data points fit a curve. • In case aggregate of SE is small, R^2 is considered ideal ($R^2 = 1$). • Inversely, if the aggregate of SE is high, then R^2 will be 0. • It is used when amount of terms are to be adjusted in a model • If inadequate terms are higher in the model, R_A^2 will decrease. • It is used as arithmetical calculation of how well a classification test appropriately recognizes the problem.
Mean Absolute Error	$MAE = \frac{1}{n} \sum_{i=1}^n x_i - \hat{x}_i $	It refers to the average of divergence among ground truth and estimated scores.	
Root Mean Square Error	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2}$	It refers to the square root of average of squared divergence among the expected and estimated score by the model.	
R^2	$coeff(R^2) = 1 - \frac{SE(\text{line})}{SE(\bar{X})}$	It measures what percentage of dissimilarity in target (X) is interpreted by the dissimilarity in regression line (Y).	
Adjusted R^2	$R_A^2 = 1 - \left[\left(\frac{m-1}{m-n-1} \right) \cdot (1 - R^2) \right]$	It is used to analyze how better the model is equipped to data points.	
Accuracy	$ACC = \frac{tp + tn}{tp + fp + tn + fn}$	It calculates if positive and negative predictions are accurately classified.	

(continued on next page)

Table 1 (continued)

Performance Metrics	Formula	Description	Application
PSNR	$PSNR = 20\log_{10}\left(\frac{l-1}{RMSE}\right)$	It is used to measure the rectification quality of images.	<ul style="list-style-type: none"> • PSNR is beneficial if comparison is made among pictures with distinct dynamic ranges.
Structural Similarity Index	$SSIM(m,n) = \frac{(2\mu_m\mu_n + a_1)(2\sigma_{mn} + a_2)}{(\mu_m^2 + \mu_n^2 + a_1)(\sigma_m^2 + \sigma_n^2 + a_2)}$	It is a calculation of similarity among 2 images.	<ul style="list-style-type: none"> • It is a human visual system (HVS) feature based metric. • The HVS executes various image refining functions that are more advanced than other models.
Gradient Similarity Metric	$GMS = \frac{2x_a x_b + m}{x_a^2 + x_b^2 + m}$	It measures local quality map generating single quality value by utilizing standard deviation.	
Universal Quality Index (UQI)	$UQI = \frac{4\sigma_{mn}\overline{mn}}{(\sigma_m^2\sigma_n^2)((\overline{m})^2 + (\overline{n})^2)}$	The UQI is given by the equation:	<ul style="list-style-type: none"> • UQI models the correlation loss, luminance distortion, and contrast distortion.
Pixel accuracy	$PA = \frac{\sum_{i=0}^M P_{ii}}{\sum_{i=0}^M \sum_{j=0}^M P_{ij}}$	It directly detects the ratio of pixels accurately identified, to the total number of pixels.	<ul style="list-style-type: none"> • It is useful in determining how accurately a segmentation model operates.
Intersection Over Union	$IoU = J(X, Y) = \frac{ X \cap Y }{ X \cup Y }$	It is characterized as the region of convergence among the predicted segmentation plan and the ground truth, to the area of union among the predicted segmentation and the ground truth.	<ul style="list-style-type: none"> • It is useful in CV applications.
Mean Pixel Accuracy	$MPA = \frac{1}{M+1} \sum_{i=0}^M \frac{P_{ii}}{\sum_{j=0}^M P_{ij}}$	It is the enhanced form of PA, where the ratio of accurate pixels is calculated in a per-class mode and later averaged over the aggregated categories.	<ul style="list-style-type: none"> • It is useful in determining how accurately a segmentation model operates.
Area Under Curve		It offers an overall assess of functioning over all possible categorization thresholds.	<ul style="list-style-type: none"> • AUC values extend from 0 to 1 and a model whose anticipation is 100% inaccurate has an AUC of 0; and whose anticipation are 100% precise has an AUC of 1.
Mean IoU		It is employed in characterizing the evaluation of contemporary segmentation techniques.	
Jaccard Index	$JA = TP / (TP + FN + FP)$	It is referred as the ratio of measure of the intersection of 2 bitsets to the measure of union.	<ul style="list-style-type: none"> • It is used for calculating correlation and dissimilarity of sample sets.
Dice	$Dice = 2 \frac{ X \cap Y }{ X \cup Y }$	It is a measure of correlation of 2 samples.	<ul style="list-style-type: none"> • It finds application in image segmentation to compare outputs of algorithms with reference masks in therapeutics.

3.4.28. Region of convergence (ROC)

It refers to a 2 D graph that predicts the tradeoff among TPR and FPR.

3.4.29. Area under curve (AUC)

It offers an overall assess of functioning over all possible categorization thresholds. AUC values extend from 0 to 1 and a model whose anticipation is 100% inaccurate has an AUC of 0; and whose anticipation are 100% precise has an AUC of 1.

3.4.30. ROC-AUC

The parameter is essentially a means of estimating the interpretation of an ML approaches. AUC calculates the capability of a binary classifier to make a distinction among categories and is employed as an outline of the ROC curve.

3.4.31. Matthews correlation coefficient (MCC)

The association among estimated class and the ground truth is referred to as MCC (equation (46)).

$$MCC = \frac{(TP \times TN - FP \times FN)}{(TP + FP)(TP + FN)(TN + FP)(TN + FN)} \quad (46)$$

3.4.32. Divergence value

The DV can be calculated by the equation (47).

$$DV = FP + FN / TP + FN \quad (47)$$

3.4.33. Jaccard index

It is referred as the ratio of measure of the intersection of 2 bitsets to the measure of union. The value of JA is computed utilizing the equation (48).

$$JA = TP / (TP + FN + FP) \quad (48)$$

3.4.34. Structural similarity index (SSIM)

It is a calculation of similarity among 2 images and is given by the equation (49):

$$SSIM(m,n) = \frac{(2\mu_m\mu_n + a_1)(2\sigma_{mn} + a_2)}{(\mu_m^2 + \mu_n^2 + a_1)(\sigma_m^2 + \sigma_n^2 + a_2)} \quad (49)$$

μ_m, μ_n = average of m and n respectively.

σ_m^2, σ_n^2 = variance of m and n respectively

σ_{mn} = covariance of m and n

Constraints $a_1 = (K1X)^2$ and $a_2 = (K2X)^2$

K = dynamic range of pixel values

3.4.35. Peak signal to noise ratio (PSNR)

It refers to the ratio of maximum probable power of signal to the potential of degrading noise which influences the accuracy of the image. It is used to measure the rectification quality of images (equation (50)).

$$PSNR = 20\log_{10}\left(\frac{l-1}{RMSE}\right) \quad (50)$$

l=number of intensity levels

5. Experimental evaluation

With the aim to evaluate the segmentation performance of the distinct methodologies, experiments are performed using 33 different segmentation techniques in this study. The tests are evaluated using MATLAB and Python 3.9.13 on windows 10 operating system. For administering varied segmentation strategies on pictures from distinct databases the EfficientNet approach is executed on aerial pictures. The implementation encompasses specific steps for data refining, constructing a model, and estimation. The diversity of the database is confined to 20 categories. The archives incorporate details as to **discrete categories of entities** which could be counted in every picture. Every

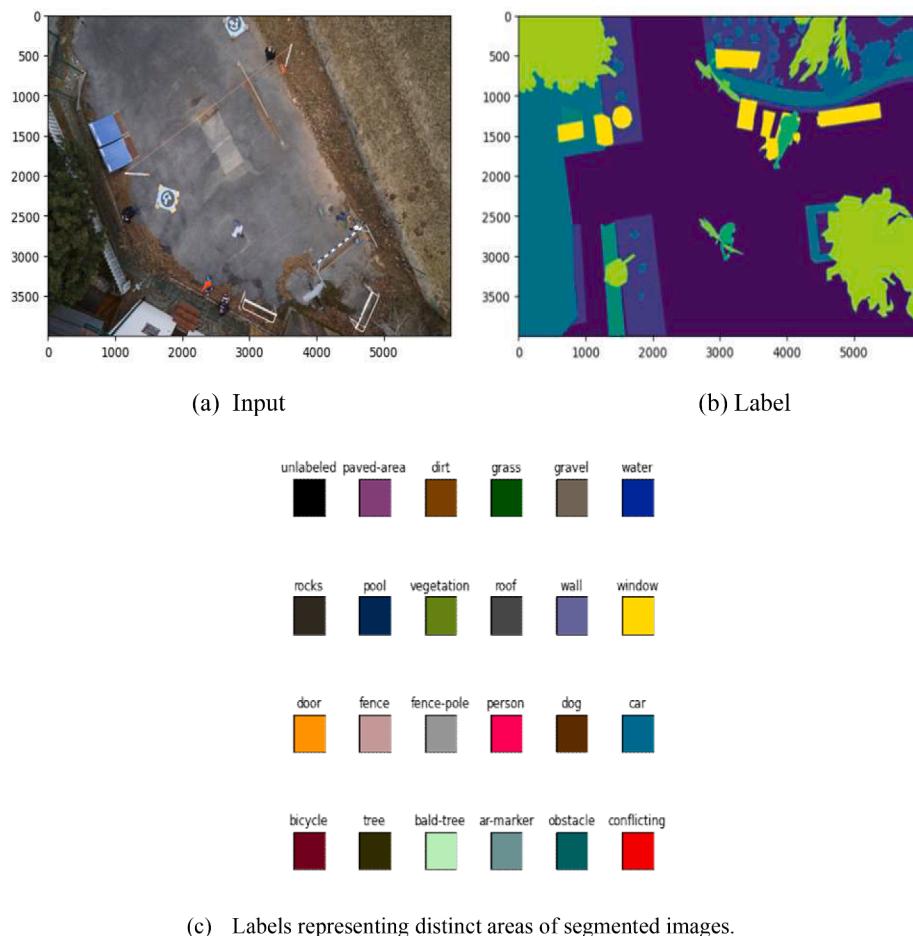


Fig. 16. Illustration of EfficientNet approach on aerial pictures (<https://towardsdatascience.com/semantic-segmentation-of-aerial-imagery-captured-by-a-drone-using-different-u-net-approaches-91e32c92803c>)

category of entity must comprise a **distinct color for discrimination of all categories**. The **fig. 16** illustrates sample of actual picture and labeled picture. **Table 2** illustrates the results for EfficientNet approach. Similarly, another segmentation methodology called SegNet has been effectuated on the same database. The **fig. 17** illustrates several samples of actual pictures, labeled pictures and resultant anticipated pictures. The **table 3** specifies the outcomes for the SegNet approach. Another approach called DeepLabV3 is executed on the same database and the outcomes are demonstrated in **table 4**. The **fig. 18** illustrates the input, output and labeled pictures respectively. Another approach called MobileUnet is executed on the same database and the outcomes are demonstrated in **table 5**.

The **fig. 19** illustrates the input, labeled and output pictures respectively. Another approach called ResNet50 is executed on the same database and the outcomes are demonstrated in **table 6**. The **fig. 20** illustrates the input, labeled and output pictures respectively. Contemplating additional database of aerial pictures accessed from MBRSC satellites and depicted by segmentation in 6 categories. An approach called InceptionResNetV2 is executed on this database after amplifying the data. The resultant outcomes have been demonstrated in **table 7**. The **fig. 21(a)** illustrates input, augmented, original mask and segmented mask pictures and **(b)** illustrates input, mask and anticipated mask pictures. Another approach called MultiUNet is executed on the same database and the outcomes are demonstrated in **table 8**. The **fig. 22** illustrates the input and output pictures respectively. A contrast analysis of 2 segmentation procedures on the same database has been executed on the same database and the results have been demonstrated in **table 9**. The **fig. 23** illustrates the input pictures and the corresponding mask,

anticipated mask and the erroneous picture for (a) MultiUNet and (b) ResNet50 respectively. The relevance of image segmentation in found in therapeutics too and the recognition of ailment gets less troublesome. For showcasing therapeutic image segmentation evaluation, an X-ray based dataset for monitoring the aggregate of pneumonia and normal pictures have been utilized which is shown in **fig 24**. The contrast of 3 segmentation methodologies has been demonstrated in **table 10**. The outcomes are distinguished on the basis of normal and pneumonia cases.

Now by utilizing the same database, another contrast-based evaluation of 3 distinct segmentation approaches have been executed. The pneumonia cases are represented by 0 while the normal cases are represented by 1 which is illustrated in **fig 25**. The outcomes for both these cases are demonstrated individually and are listed in **table 11**. In subsequent example an attention approach for detecting COVID is executed. The outcomes are demonstrated in **table 12** and **fig. 26** illustrates the resultant pictures. A distinct approach for identifying different eye diseases as illustrated in **fig 27** manoeuvring EfficientNet method is discussed in this example. **Table 13** demonstrates outcomes for segmentation technique utilized. Identification of skin abrasion aiding segmentation approaches is another crucial aspect. In this part we maneuverer ISIC database exhibiting the aggregate count of pictures of every category and executing the DenseNet121 methodology. The **fig. 28** illustrates the aggregate count of all the pictures prior augmentation. **Table 14** demonstrates the evaluation outcomes. In distinct example utilizing the same database, the contrast amongst 3 segmentation approaches is demonstrated in **table 15**. The contrast amongst various segmentation models for skin cancer analysis has been substantiated and the database utilized here is HAM10000. The **fig. 29**

Table 2
Results for EfficientNet segmentation

Evaluation Metric	EfficientNet
Loss	0.09
Accuracy	1.0
F-1	1.0
AUC	1.0
Validation loss	0.51
Validation Accuracy	0.87
Validation F-1	0.88
Validation AUC	0.93

illustrates the 7 determination codes in the database: marked 0 to 7 for specific abrasion category. Table 16 demonstrates evaluation of 3 segmentation methodologies. The PH2 dataset illustrations have been utilized for performing otsu and region growing based segmentation procedures and the results are evaluated on the basis of dice, recall, f-1, accuracy, SE, SP, precision, JA, DV. The fig. 30 (a) illustrates input images, (b) otsu segmented resultant and (c) region growing resultant. Tables 17-18 exhibit otsu's and region growing segmentation outcomes on distinct classifiers. The subsequent case exhibits the mechanism where regions can be determined from the therapeutic pictures (Brain MRI) wherein a choice is made if the growth is carcinogenic or not in the therapeutic pictures. The DenseNet121 approach is executed for this example and the outcomes have been illustrated in table 19. Fig. 31 illustrates the sample pictures after pre-refining. In distinct example the sentiment realization from face expressions accommodating 7 divergent emotions is discussed. Each expression is labelled with numerical values 0-6 which has been illustrated in fig 32. The result for VGG19 approach is discussed in table 20. Fig. 33 illustrates the resultant predicted output pictures. In a subsequent example discernment of pictures involving automobile (represented by 1) or not (represented by 0) utilizing 3 segmentation approaches is executed. Fig. 34(a) illustrates sample

Table 3
The results for SegNet based segmentation.

Evaluation Metric	SegNet
Loss	1.50
Accuracy	0.55

Table 4
Results for DeepLavV3 approach on aerial pictures

Evaluation Metric	DeepLabV3
Loss	0.07
Mean IoU	0.66
Validation loss	0.18
Validation mean IoU	0.49

pictures for both classes while fig. 34(b) illustrates the test results. Table 21 demonstrates the outcomes for distinct segmentation models. Considering a distinct database i.e., ‘cityscapes’ contrast-based evaluation of 9 distinct segmentation approaches has been executed. The resultant pictures for each segmentation methodology have been illustrated in fig. 35(a)-(i). The Table 22 demonstrates the evaluation performances of all segmentation methodologies. The standard test images ‘Lenna’ and ‘Mandrill’ have been utilized to carry out thresholding-based segmentation approaches and results are evaluated relying on PSNR and SSIM values. The resultant threshold pictures are illustrated in figs. 37 and 38 and Table 24, 25, 26 and 27 depicts contrast of PSNR and SSIM values amongst thresholding-based segmentation approaches for test pictures lenna and mandrill. A set of 200 original photos from the PH2 dataset were used to evaluate the segmentation methods recently outlined; the qualitative findings are displayed below. Fig. 36 (i-vii) shows the segmentation for several techniques, with the representation based on: (i) the original image; (ii) Otsu; (iii-v) Multi-level otsu; (vi-vii) Niblack and sauvola thresholding; (vi) active contour; (vii)watershed;

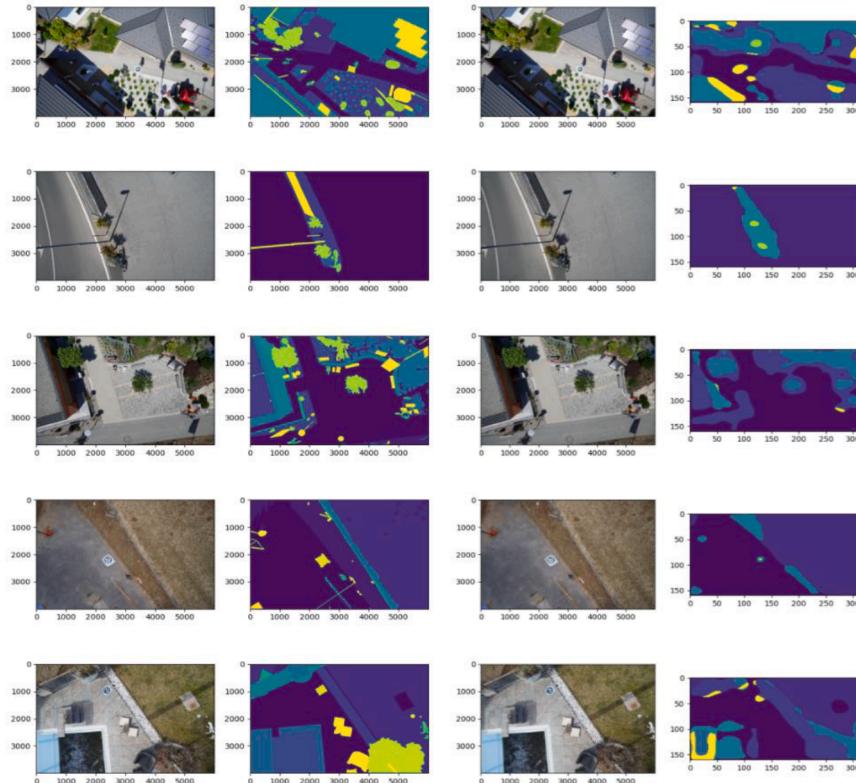


Fig. 17. Illustration of several samples of actual pictures, labelled pictures, and resultant predicted pictures (<https://www.kaggle.com/datasets/bulentsiyah/semtic-drone-dataset/code>)

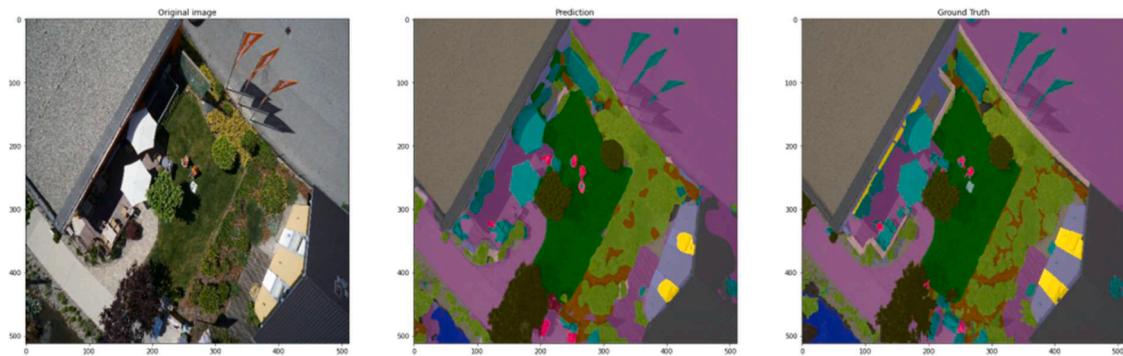


Fig. 18. Illustrates the input, output and labeled pictures respectively for DeepLabV3 (<https://www.kaggle.com/datasets/bulentsiyah/semantic-drone-dataset/code>)

Table 5
Results for MobileUNet based segmentation approach on aerial pictures.

Evaluation Metric	MobileUNet
Loss	0.71
Accuracy	0.81
Validation loss	0.67
Validation Accuracy	0.81

(viii) FMM; (ix)k-means clustering and (vii) FCM. From the 200 photos in the PH2 database, we have gathered the skin cancer photographs. Each image had a resolution of 768*560, and they were categorized into three classes: common nevi, atypical nevus, and melanoma. There are 40, 80, and 80 photos in each of the ensuing classes, respectively. We use the textural characteristics of skin cancer images to verify the accuracy and efficacy of our proposed technique. **Table 23** displays the segmentation scores for each method.

6. Principles, advantages and disadvantages of traditional image segmentation methods

Generally speaking, there has been a shift in image segmentation progress from single-image orientation to segmentation that utilizes common features of vast data, from coarse-grained to fine-grained, and from manual feature extraction to adaptive learning. The variety of picture types being captured by image acquisition technologies is increasing the issues associated with segmenting images of varying sizes,

Table 6
Results for ResNet50 approach on aerial pictures

Evaluation Metric	ResNet50
Loss	0.14
Accuracy	0.95

resolutions, and imaging modalities. Researchers anticipate using a general network that is more flexible and capable of generalization. Deep NN research has demonstrated clear benefits in object recognition and scene interpretation since the FCN was suggested. Subsequent research endeavours will continue to center around deep NN, with the goal of enhancing the network's precision, reactivity, and resilience. Image segmentation moved from the CNN phase to the transformer generation in 2021 thanks to the significant advancement in CV research produced by the Swin Transformer. This development could lead to further advancements in the domain of computer vision analysis.

1. DL does, however, have several drawbacks. For example, its inexplicability restricts the resilience, dependability, and performance optimization of its subsequent duties. The proliferation of vision-based technologies has been aided by the quick development of deep learning (DL) and improvements in device capabilities like as memory size, processing speed, energy use, optics, and image sensor resolution. These advancements have also increased performance and reduced costs. DL is sometimes unnecessary since standard image processing can frequently tackle a given challenge more

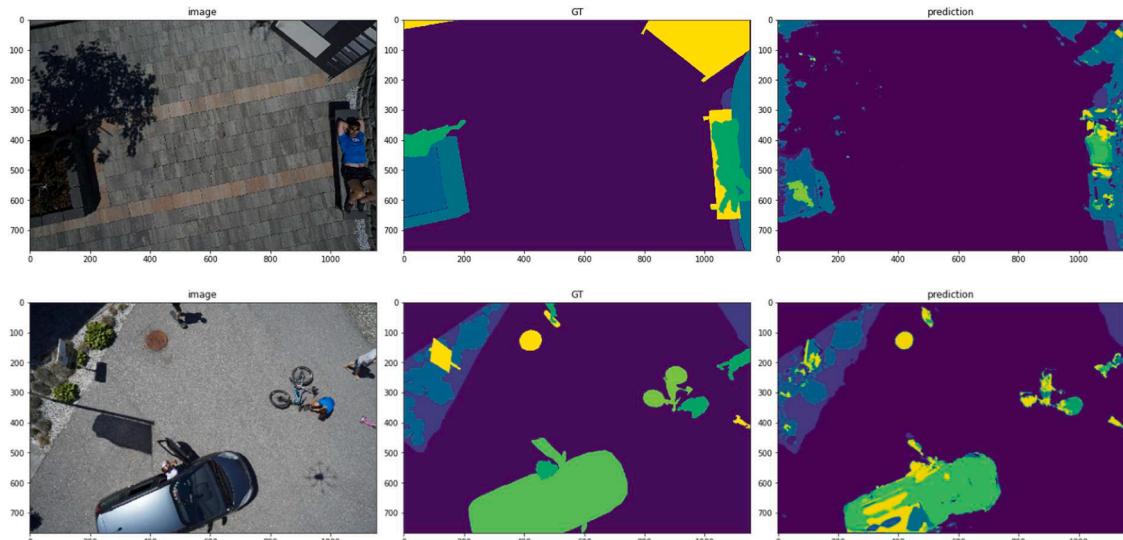


Fig. 19. Illustration of the input, labelled and output pictures respectively for MobileUnet (<https://www.kaggle.com/datasets/bulentsiyah/semantic-drone-data-code>)

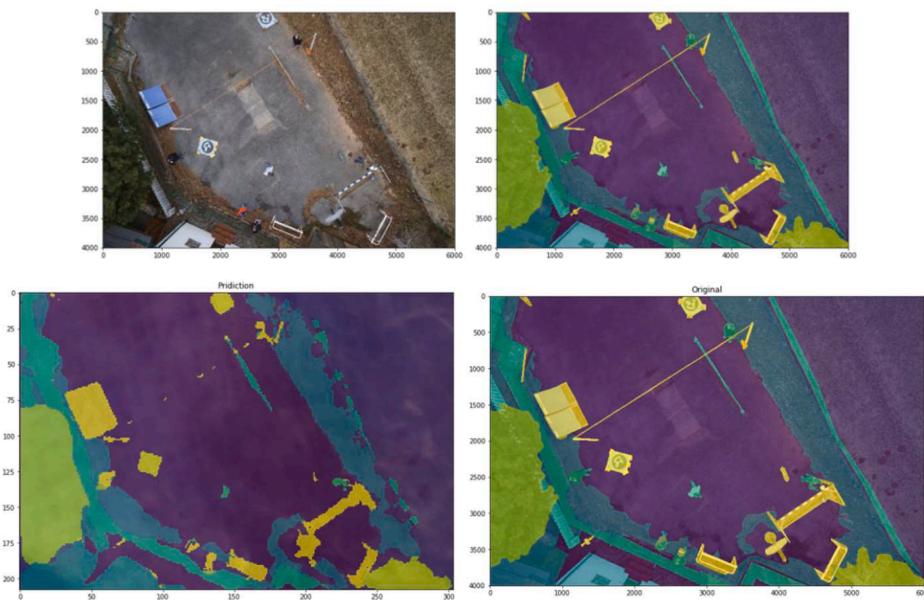


Fig. 20. (a). Illustration of input and label pictures (b). Illustration of the predicted mask and input label for ResNet50 (<https://www.kaggle.com/datasets/bulent-siyah/semantic-drone-dataset/code>)

Table 7
Results of InceptionResNetV2 approach on aerial pictures

Evaluation Metric	InceptionResNetV2
Loss	0.24
Dice	0.85
Accuracy	0.91
Validation loss	0.48
Validation dice	0.79
Validation Accuracy	0.84

Table 8
Results for MultiUNet approach on aerial pictures

Evaluation Metric	Multi UNet
Loss	0.92
Accuracy	0.82
Jaccard	0.64
Validation loss	0.93
Validation accuracy	0.80
Validation Jaccard	0.60

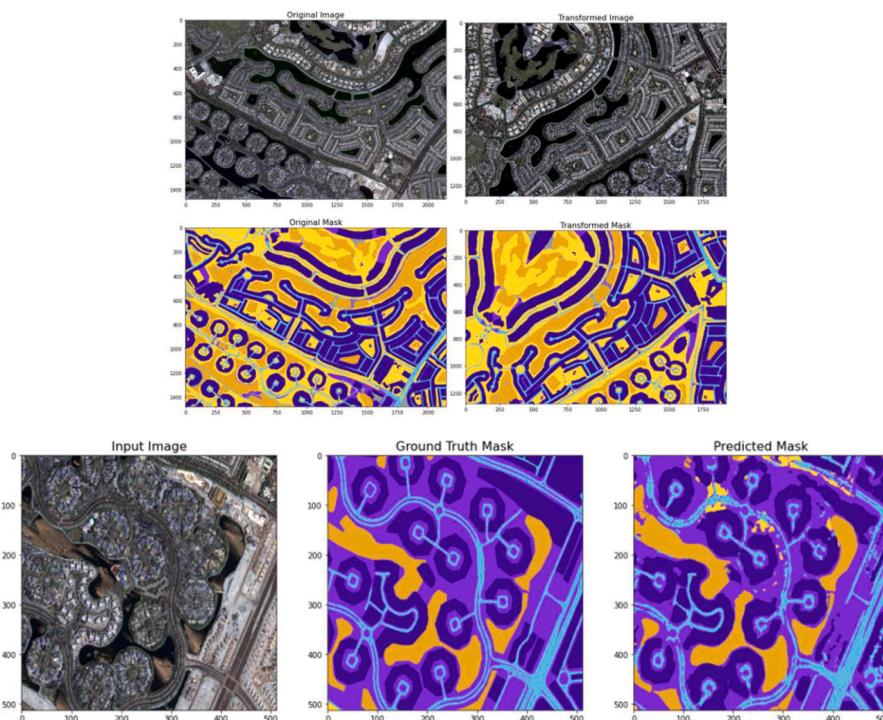


Fig. 21. (a). Illustration of input, augmented, original mask and segmented mask pictures (b). Illustration of input, mask and anticipated mask pictures (<https://www.kaggle.com/datasets/humansintheloop/semantic-segmentation-of-aerial-imagery>)

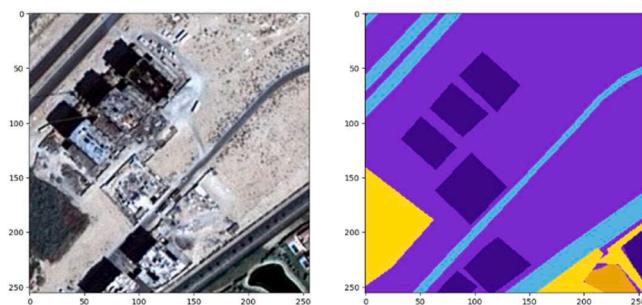


Fig. 22. Illustration of the input and output pictures respectively (<https://www.kaggle.com/datasets/humansintheloop/semantic-segmentation-of-aerial-imagery>)

Table 9
Comparative analysis of MultiUNet and ResNet50 segmentation approaches

Evaluation Metric	Multi UNet	ResNet 50
Loss	0.10	0.04
Accuracy	0.95	0.98
Mean IoU	0.39	0.62
Validation loss	0.61	0.93
Validation accuracy	0.86	0.86
Validation mean IoU	0.38	0.54

accurately and with less lines of code than deep learning. A deep neural network's learned characteristics are unique to its training dataset, therefore if the dataset is poorly designed, the network is unlikely to function well on images that are not part of the training set. Conversely, methods like as SIFT and even basic color thresholding and pixel counting techniques are quite broad and work in a similar way for any picture; they are not class-specific.

The following are the prevailing image segmentation study recommendations and issues: The areas of image segmentation analysis that remain hotspots include semantic segmentation, instance segmentation, and panoramic segmentation. While panoramic segmentation combines

semantic and instance segmentation, it gives every pixel in the picture a class name and a segment ID. In instance segmentation determines the pixel portions encompassed in every instance. Building an efficient network that consistently detect both huge inter-category variations and small intra-category variations is a demanding task, particularly for panoramic segmentation where countable or uncountable examples are challenging to distinguish in a single workflow;

2. As image capture tools (such LiDAR cameras) have become more widely used, study areas such as RGB-depth, 3D-point clouds, voxels, and mesh segmentation have grown. These areas have a wide range of applications, including identification of faces [95], self-driving vehicles, VR, AR, architectural modeling, etc. While some advancements have been made in the field of 3D image segmentation research—such as the use of machine learning algorithms like SVM, random forest, and AdaBoost, and traditional algorithms like region growth, random walks, and clustering—the depiction and computation of 3D data—which is unstructured, ineffective, cluttered and haphazardly distributed—remains a significant challenge;

3. Lack of fine-grained descriptors or datasets makes it challenging to train the network using supervised learning algorithms in several domains. In these situations, the network can be trained on the reference set first, allowing the smaller-scale variables to be fixed. The fully connected layer or a subset of the high-level variables is then used for training on the small-sample information set. Semi-supervised and unsupervised semantic segmentation can be chosen in these scenarios. Transfer learning is this, and it doesn't need a lot of labelled samples. Another option is reinforcement learning; however this approach is not often investigated in the context of image segmentation. Additionally, a popular area of investigation is few-shot imaging semantic segmentation.

4. The fact that deep learning networks need a lot of processing power during training further demonstrates the deep neural network's computational complexity. In some domains, such as video processing, real-time (or almost real-time) segmentation is necessary to satisfy the minimum 25 frames per second required by human vision, although the majority of networks in use today fall well short of this benchmark. There is still much space for enhancement in the

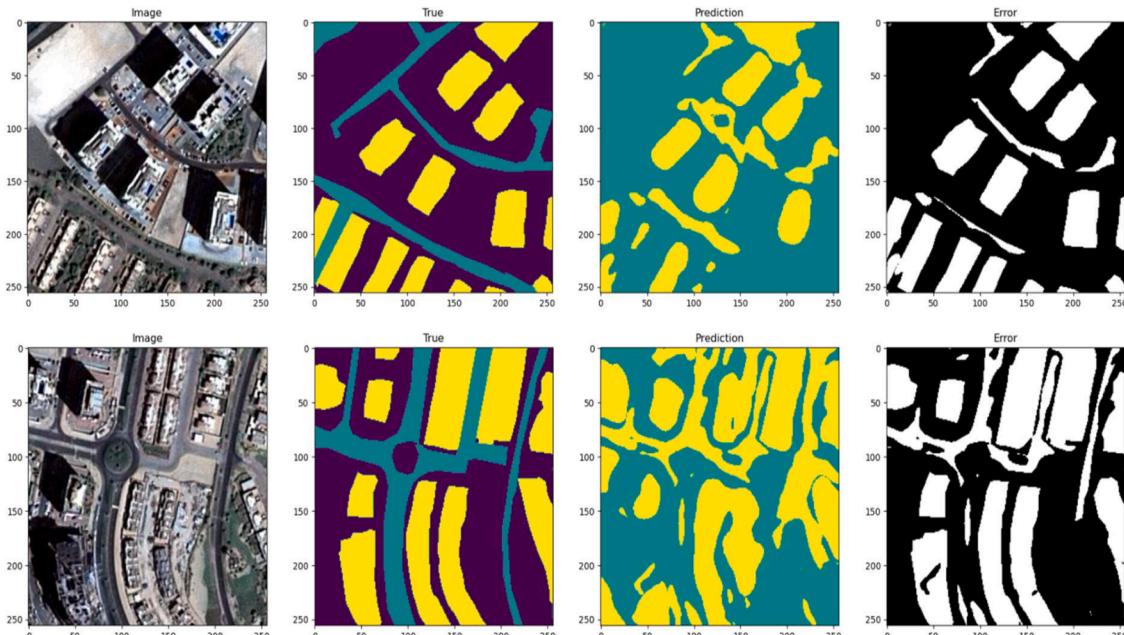


Fig. 23. (a). Illustration of the input pictures and the corresponding mask, anticipated mask and the erroneous picture for MultiUNet **(b).** Illustration of the input pictures and the corresponding mask, anticipated mask and the erroneous picture for ResNet50 (<https://www.kaggle.com/datasets/humansintheloop/semantic-segmentation-of-aerial-imagery>).

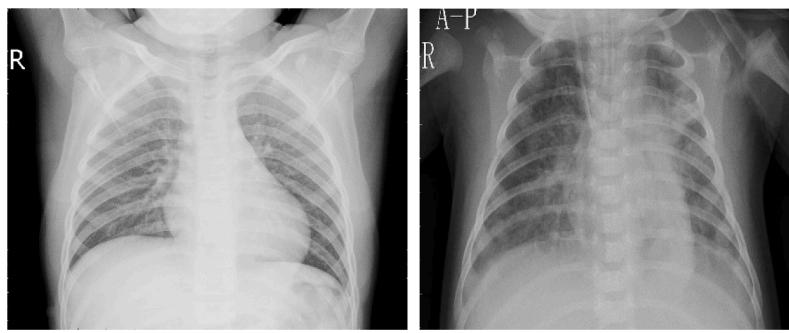


Fig. 24. Illustration of normal and pneumonia X-ray pictures (<https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia/code>)

Table 10
The contrast of AlexNet, ResNet152 and ResNet50 methodologies

Evaluation Metric	AlexNet (Normal case)	AlexNet (Pneumonia)	ResNet152 (Normal)	ResNet152 (Pneumonia)	ResNet50 (Normal)	ResNet50 (Pneumonia)
Loss	0.03	0.03	0.16	0.16	0.29	0.29
Accuracy	0.96	0.96	0.94	0.94	0.87	0.87
Validation loss	0.16	0.16	0.16	0.16	0.25	0.25
Validation accuracy	0.94	0.94	0.94	0.94	0.91	0.91
Precision	0.93	0.89	0.92	0.80	0.87	0.77
Recall	0.79	0.97	0.61	0.97	0.53	0.95
F1	0.86	0.93	0.73	0.88	0.66	0.85

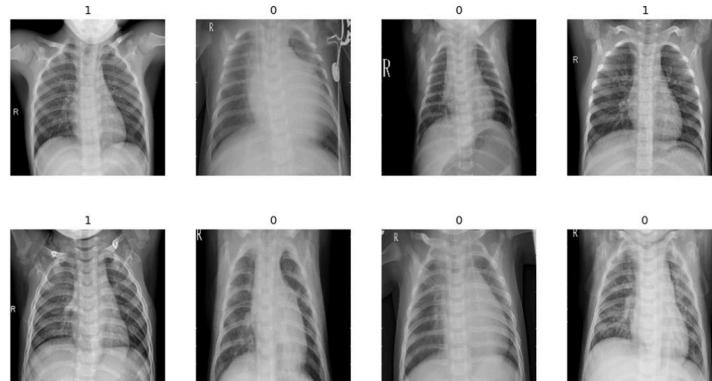


Fig. 25. Illustration of the pneumonia cases denoted by 0 while the normal cases by 1(<https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia/code>)

Table 11
Results of VGG16, AlexNet and MobileNetV3 for pneumonia identification

Evaluation Metric	VGG16 (Normal case)	AlexNet (Pneumonia)	MobileNetV3 (Normal)	MobileNetV3 (Pneumonia)	ResNet50 (Normal)	ResNet50 (Pneumonia)
Loss	0.12	0.12	0.20	0.20	0.23	0.23
Accuracy	0.95	0.95	0.91	0.91	0.90	0.90
Validation loss	0.13	0.13	0.15	0.15	0.23	0.23
Validation accuracy	0.95	0.95	0.93	0.98	0.90	0.90
Precision	0.83	0.99	0.82	0.80	0.81	0.98
Recall	0.98	0.92	0.96	0.91	0.95	0.91
F1	0.90	0.95	0.88	0.95	0.87	0.94

trade-off between real-time efficiency and model correctness, even if some lightweight models have somewhat accelerated the segmentation process.

Some of the effective implementation of segmentation approaches are:

6.1. Content-based image retrieval

This is also called as query by image content (QBIC) and its relevance is found in CV approaches to the digital image refining issue. ‘Content-based’ signifies that the search examines the constituents of the image instead of the data elements like descriptors, labels, or depictions clearly

Table 12

Results of attention UNet for COVID identification

Evaluation Metric	Attention UNet
Loss	0.17
Dice	0.82
Sensitivity	0.75
Specificity	0.99
Validation loss	0.26
Validation dice	0.72
Validation Sensitivity	0.78
Validation Specificity	0.99
Binary Accuracy	0.99
Validation Binary accuracy	0.99

related to the image. The expression ‘content’ indicates colours, shapes, textures, or all additional particulars which can be drawn from the image. QBIC is beneficial because studies relying on data elements

Table 13

Results for Attention UNet technique for eye disease recognition

Evaluation Metric	Attention UNet
Loss	0.33
Accuracy	0.89
Validation loss	0.29
Validation Accuracy	0.91

depend on interpretation quality and integrity. The significance of CBIR has expanded because of the drawbacks implicit in data frameworks, and the wide range of potential utilizations for effective image retrieval. Textual details about images can be investigated manoeuvring contemporary mechanisms, but this involves humans to characterize every image in the dataset. It can be unrealistic for huge datasets or for images which are produced automatically. It is also likely to overlook

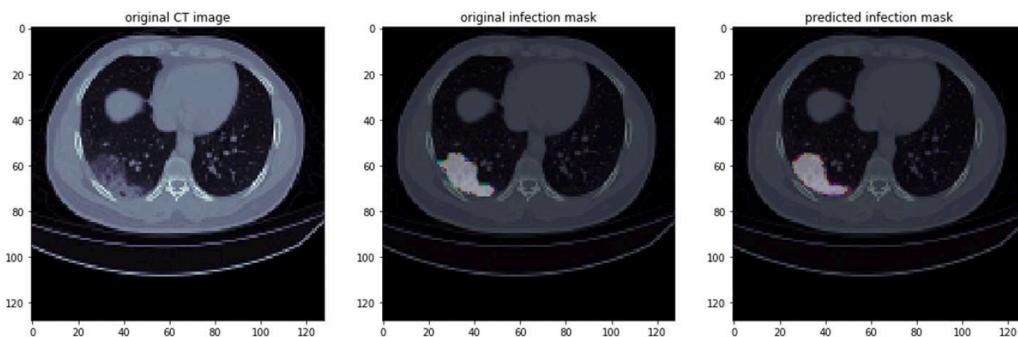


Fig. 26. Illustration of the original picture, original contaminated mask and output mask respectively (<https://www.kaggle.com/datasets/andrewmvd/covid19-ct-scans>)

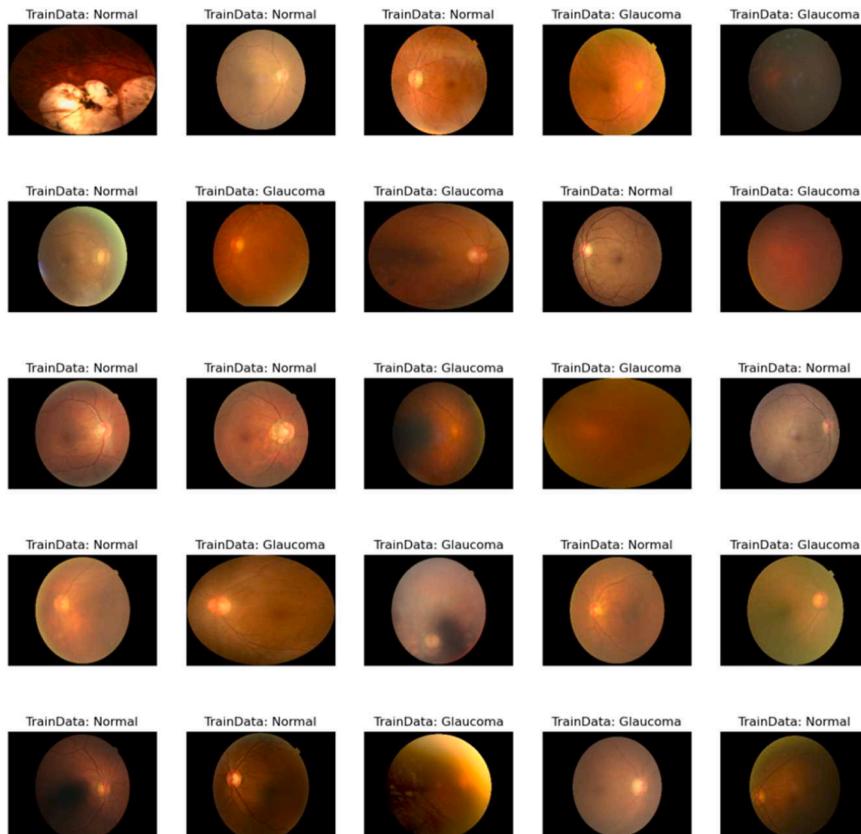


Fig. 27. Illustration of the distinct eye disease cases (<https://www.kaggle.com/code/faizalkarim/pytorch-eye-disease-classification-93-7>)

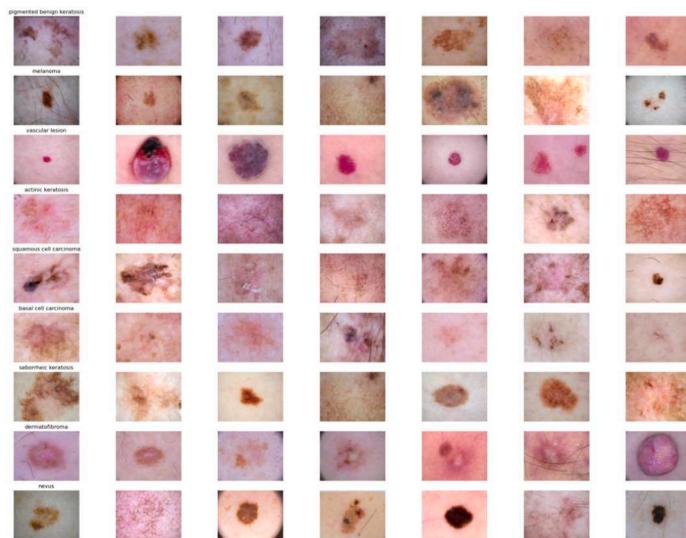


Fig. 28. Illustration of the aggregate count of all the ISIC pictures prior augmentation (<https://www.kaggle.com/code/blamerx/skin-cancer-classification-densenet-121-and-aug>)

Table 14
Results for DenseNet121 approach for skin cancer recognition

Evaluation Metric	DenseNet121
Loss	0.47
Test Accuracy	0.91
Precision	0.91
Recall	0.91
F1	0.91
Cohen Kappa	0.90

Table 15
Results for CNN, ResNetV2 and VGG16 approaches for skin cancer recognition
(<https://www.kaggle.com/datasets/nodoubtome/skin-cancer9-classesisic/code>)

Evaluation Metrics	CNN	ResNetV2	VGG16
Loss	0.35	0.02	0.68
Accuracy	0.83	0.99	0.54
Validation loss	0.31	3.10	0.68
Validation Accuracy	0.83	0.49	0.54
Test loss	0.31	1.55	0.68
Test Accuracy	0.83	0.72	0.54

images which employ diverse synonyms in their depictions. Methodologies relying on classifying images in semantic categories can avoid the misclassification issue but will need more attempt by an employer to discover images. Various standards have been implemented to classify images, but these face computing and mis-classification drawbacks. Preliminary QBIC frameworks were implemented to discover datasets relying upon image colour, texture, and shape traits. After these frameworks were implemented, the urge for convenient divisions became evident. As a result, attempts in the QBIC domain initiated to incorporate human-composed plan that tested to respond to the demand of the employer operating the search. This simply means incorporation of uncertainty approaches that might permit graphic semantics, queries which might incorporate user response, procedures that might comprehend machine learning, and networks which may recognize

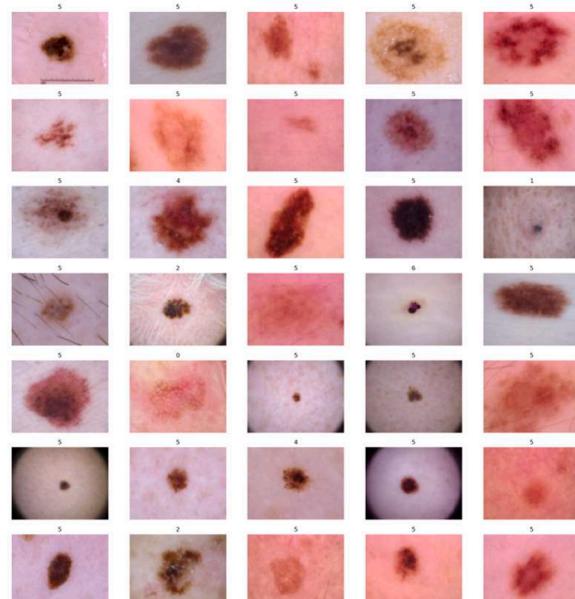


Fig. 29. Illustration of the 7 determination codes in the database: marked 0 to 7 for specific abrasion category (<https://www.kaggle.com/datasets/surajghuwalewala/ham1000-segmentation-and-classification/code>)

employer satisfaction extents. A simple explanation of it is provided in fig. 39.

6.2. Machine vision

Machine vision as an automatic control engineering exercise can be examined from CV, a sort of computer science. It seeks to combine contemporary methodologies in novel manners and evaluate them to unravel real world issues. The title is the predominant one for the

Table 16
Results for denseNet121, ResNet50 and InceptionV3

Evaluation Metrics	DenseNet121	ResNet50	InceptionV3	
Loss	0.03	0.04	1.0	
Accuracy	0.99	0.99	0.68	
Validation loss	0.91	0.86	1.28	
Validation Accuracy	0.81	0.78	0.64	
Precision	0 1 2 3 4 5 6	0.60 0.81 0.61 0.95 0.58 0.91 0.88	0.45 0.78 0.64 0.91 0.63 0.88 0.92	0.0 0.0 0.0 0.0 0.0 0.65 0.0
Recall	0 1 2 3 4 5 6	0.59 0.48 0.66 0.80 0.69 0.91 0.74	0.69 0.67 0.54 0.40 0.47 0.94 0.74	0.0 0.0 0.0 0.0 0.0 1.0 0.0
F1	0 1 2 3 4 5 6	0.59 0.60 0.63 0.87 0.63 0.91 0.81	0.54 0.72 0.58 0.56 0.54 0.91 0.82	0.0 0.0 0.0 0.0 0.0 0.79 0.0

exercises in manufacturing computerized environments and is also utilized for these functions in distinct domains. The comprehensive machine vision procedure incorporates organizing the particulars of the regulations and project, and then generating an outcome. In contemplation of run-time, the procedure initiates with imaging, accompanied by automated scrutiny of the image and withdrawal of the necessitated information. DL training and conclusion inflict higher refining execution prerequisites. Different levels of refining are typically employed in a progression that concludes the preferred outcome. A conventional cycle might initiate with refining tools like filters that transforms the image, accompanied by withdrawal of characteristics, then withdrawal of information from such entities, and then passing that information, or contrasting it with target scores to generate and display "pass/fail" outcomes. Machine vision image refining processes incorporates:

ü Registration: Fusion of adjoining 2D or 3D images

Filtering

ü Thresholding: It initiates with defining a gray level which will be beneficial for the subsequent steps. This score is later utilized to split parts of an illustration, and periodically to modify very part of the illustration to black and based on below or above grayscale levels.

ü Pixel count: Calculates the amount of light or dark pixels.

ü Segmentation: Segmenting a digital picture to numerous portions to facilitate and transform the illustration of an image to something which is more significant and better to scrutinize.

ü Edge detection

ü Colour Analysis

ü Blob identification and withdrawal.

ü DL processing

ü Pattern recognition.

ü Optical character recognition

ü Contrast against expected values to identify an outcome.

ü Vision based robot supervision.

6.3. Medical imaging

Medical imaging is an approach of imaging the internal parts of a body for impersonal investigation and therapeutic arbitration, and graphical display of the operation of few anatomical components or tissues. Medical imaging anticipates displaying interior formations obscured by the skin and bones, and to recognize and handle disease. It also determines a dataset of ordinary anatomy and biology to make it possible to discover malformations. The assessment and transcription approaches which are not ideally fabricated to generate images, like electroencephalography (EEG), magnetoencephalography (MEG), electrocardiography (ECG), portray other technologies generating data vulnerable to data as a variable plot versus time which comprise information about the magnitude positions. In short, such technologies are evaluated types of medical imaging in another domain of medical instrumentation. Medical imaging is observed to indicate the variety of approaches which noninvasively generate images of internal organs and can be observed as the explication to analytical inverse problems. It simply signifies the origin is reckoned from strike. In reference to prognostication radiography, the tube employs X-ray excretion that is assimilated at diverse proportions by distinct tissue sets like bone and muscle. Many of the methodologies generated for medical imaging also have technical and commercial implementations:

ü Radiography: The imaging configuration employs a large beam of X rays for illustration procuring and is the primary imaging approach accessible in advanced medicine.

ü Magnetic Resonance Imaging (MRI): It transmits a radio frequency (RF) impulse at the vibrant amount of the hydrogen particles on H₂O



(a)



(b)



(c)

Fig. 30. (a). Original images from PH2 dataset (b) Otsu Segmented images (c) Region Growing segmentation images

Table 17

Results for Otsu based segmentation on PH2 dataset on different classifiers.

Classifier	Sensitivity	Specificity	Accuracy	Precision	Recall	F-measure	DI	JA	DV
Complex Tree	79.74%	9.52%	65%	76.82%	79.74%	77.41%	78.26%	64.28%	44.30%
Medium Tree	79.74%	9.52%	65%	76.82%	79.74%	77.41%	78.26%	64.28%	44.30%
Simple Tree	80.68%	9.09%	72%	86.58%	80.68%	81.25%	83.52%	71.71%	31.81%
Linear SVM	82%	0/0	82%	100%	82%	81.25%	90.10%	82%	18%
Quadratic SVM	81.39%	14.82%	72%	85.36%	81.39%	82.53%	83.33%	71.42%	32.55%
Cubic SVM	82.71%	21.05%	71%	81.70%	82.71%	80.98%	82.20%	69.79%	35.80%
Fine Gaussian SVM	81.81%	0/10	81%	98.78%	81.81%	88.26%	89.50%	81%	19.19%
Medium Gaussian SVM	82%	0/0	82%	100%	82%	90.10%	90.10%	82%	18%
Coarse Gaussian SVM	82%	0/0	82%	100%	82%	90.10%	90.10%	82%	18%
Fine KNN	82.27%	19.04%	69%	79.26%	82.27%	80.12%	80.74%	67.70%	39.24%
Medium KNN	81.63%	0/2	80%	97.56%	81.63%	88.23%	88.88%	80%	20.40%
Coarse KNN	82%	0/0	82%	100%	82%	90.10%	90.10%	82%	18%
Cosine KNN	82%	0/0	82%	100%	82%	90.10%	90.10%	82%	18%
Cubic KNN	81.63%	0/2	80%	97.56%	81.63%	88.23%	88.88%	80%	20.40%
Weighted KNN	84.61%	44.44%	81%	93.90%	84.61%	88.13%	89.01%	80.20%	20.87%
Boosted Trees	82%	0/0	82%	100%	82%	90.10%	90.01%	82%	18%
Bagged Trees	81.25%	0/4	78%	95.12%	81.25%	88.23%	87.64%	78%	22.91%
Subspace Discriminant	80.85%	0/6	76%	92.68%	80.85%	82.35%	86.36%	76%	25.53%
Subspace KNN	82.02%	18.18%	75%	89.02%	82.02%	84.79%	85.38%	74.48%	28.08%
RUSBoosted Trees	75%	13.33%	38%	36.58%	75%	48.64%	49.18%	32.60%	1.55%

Table 18

Region growing based segmentation results on PH2 dataset on different classifiers.

Classifier	Sensitivity	Specificity	accuracy	Precision	recall	f-measure	DI	JA	DV
Complex Tree	83.95%	23.31%	73%	82.92%	83.95%	82.42%	83.43%	71.57%	33.33%
Medium Tree	83.95%	26.31%	73%	82.92%	83.95%	82.42%	83.43%	71.57%	33.33%
Simple Tree	82.75%	23.07%	75%	87.80%	82.75%	84.02%	85.20%	74.22%	28.73%
Linear SVM	83.50%	66.66%	83%	98.78%	83.50%	89.50%	90.50%	82.65%	17.52%
Quadratic SVM	85.39%	45.45%	81%	92.68%	85.39%	88.13%	88.88%	80%	21.34%
Cubic SVM	85.36%	33.33%	76%	85.36%	85.36%	84.70%	85.36%	74.46%	29.26%
Fine Gaussian SVM	82%	0/0	82%	100%	82%	90.10%	90.10%	82%	18%
Medium Gaussian SVM	82%	0/0	82%	100%	82%	90.10%	90.10%	82%	18%
Coarse Gaussian SVM	82%	0/0	82%	100%	82%	90.10%	90.10%	82%	18%
Fine KNN	84.14%	27.77%	74%	84.14%	84.14%	83.92%	84.14%	72.63%	31.70%
Medium KNN	82%	0/0	82%	100%	82%	90.10%	90.10%	82%	18%
Coarse KNN	82%	0/0	82%	100%	82%	90.10%	90.10%	82%	18%
Cosine KNN	81.44%	0/3	79%	96.34%	81.44%	87.57%	89.77%	79%	21.64%
Cubic KNN	82%	0/0	82%	100%	82%	90.10%	90.10%	82%	18%
Weighted KNN	81.63%	0/2	80%	97.56%	81.63%	88.20%	89.88%	80%	20.40%
Boosted Trees	82.10%	20%	79%	95.12%	82.10%	87.57%	90.10%	78.78%	22.10%
Bagged Trees	84.37%	75%	84%	98.78%	84.37%	90.10%	91.52%	83.50%	16.66%
Subspace Discriminant	86.17%	83.33%	86%	98.78%	86.17%	91.30%	92.57%	85.26%	14.89%
Subspace KNN	82.35%	20%	73%	85.36%	82.35%	83.23%	90.32%	72.16%	31.76%
RUSBoosted Trees	85.96%	22.78%	59%	59.75%	85.96%	69.65%	92.45%	54.44%	71.92%

Table 19

Results for DenseNet121 approach for brain tumor recognition

Evaluation Metric	DenseNet121 (Training)	DenseNet121 (Validation)
Loss	0.91	0.97
Accuracy	0.65	0.80
Precision	0.65	0.80
Recall	0.65	0.80

particles. RF projections transmit the impulse to the targeted organ. Also, the impulse is assimilated by molecules, engendering the supervision relating to the principal magnetic field to transform. By the time impulse gets shut down, the molecules "relax" back to coalition accommodating the principal magnet and radiate the waves in the procedure. The RF ejection from the hydrogen particles on H₂O is what is discerned and reorganized into the representation.

ü Nuclear Medicine: This comprises both analytical imaging and processing of ailment and indicated to as atomic drug. It permits

evaluation of biology and this task-based approximate to medical estimation has valuable utilizations in certain domains.

ü Ultrasound: This employs acute broadband acoustic waves in the megahertz range which are cast back by tissue to differing angles to generate images.

ü Elastography: This is a comparatively novel imaging configuration which plots elastic features of soft tissue. This configuration materialized in the past 2 decades. It is beneficial in medical analyses, as elasticity can distinguish normal and abnormal tissue for certain metastasis.

ü Photo acoustic Imaging: This is a contemporary compound biomedical imaging method. It integrates the benefits of visual assimilation dissimilarity by an ultrasonic dimensional persistence for deep imaging in dispersive reign.

ü Tomography: This is the imaging by segments or segmenting. The sub types of it include X-ray CT that conventionally generates a 2D illustration of the edifices. Here, a radiation of X-rays rotates across an entity under analysis apprehended by responsive effusion sensors

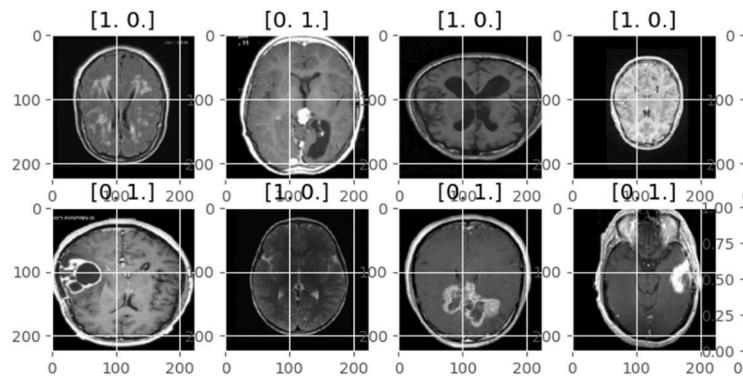


Fig. 31. Illustrations of brain MRI pictures representing carcinogenic and normal cases (<https://www.kaggle.com/code/kaledhoshme/visual-explanations-gradcam-gradcam>)



Fig. 32. Illustrations of distinct face expressions (<https://www.kaggle.com/code/enesztrk/facial-emotion-recognition-vgg19-fer2013>)

after having permeated the entity from various degrees. Later, a computer investigates the detail attained out of the recorder's sensors generating a precise representation of the entity and corresponding constituents manoeuvring the analytical standards exhibited in the Radon transform.

ü Functional near-infrared spectroscopy: It is extensively employed for brain imaging procedure.

6.4. Object detection

This is a computer mechanization referring to CV and image refining

that engages identifying circumstances of semantic entities of a particular category in digital images and videos. Considerable surveyed fields of entity identification involve face recognition and pedestrian identification. Object recognition has utilizations in several domains of CV, comprising image restoration and video surveillance.

7. Interpretation and scope

It is evident that segmenting the data has improved exceptionally from DL however several issues might impend. For example, a considerable amount of progression has been made in the therapeutic image

Table 20

Results for VGG19 approach for recognizing sentiments from face expressions.

Evaluation Metrics		VGG19
Loss		0.62
Accuracy		0.77
Validation loss		0.94
Validation Accuracy		0.68
Precision	0	0.63
	1	0.69
	2	0.52
	3	0.90
	4	0.65
	5	0.78
	6	0.56
Recall	0	0.62
	1	0.56
	2	0.52
	3	0.86
	4	0.41
	5	0.80
	6	0.80
F1	0	0.63
	1	0.62
	2	0.52
	3	0.88
	4	0.50
	5	0.79
	6	0.66

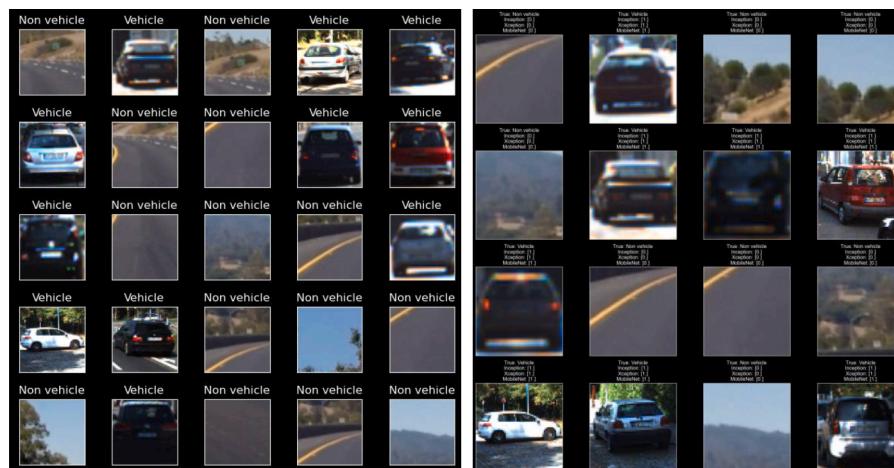
partitioning experimentation, yet the influence cannot respond to the demands of empirical implementations. The root cause of this issue is that the contemporary therapeutic image segmentation investigations possess certain complications and troubles. Clinical therapeutic diagnostics circumstances are convoluted and assorted but artificial intelligence (AI) analysts do not comprehend clinical requirements. Pathologists do not interpret the mechanism of AI. Consequently, AI

Table 21

Contrast analysis of Inception, Xception and MobileNet techniques for automobile recognition

Evaluation Metrics	Inception	Xception	MobileNet
Loss	0.05	0.08	0.44
Accuracy	0.99	0.99	0.80
Validation loss	0.09	0.01	0.19
Validation Accuracy	1.00	1.00	1.00
Test Loss	0.03	0.03	0.08
Test Accuracy	1.00	1.0	1.0
AUROC	1.0	1.0	1.0
Precision	0	1.00	1.00
	1	1.00	1.00
Recall	0	1.00	1.00
	1	1.00	1.00
F1	0	1.00	1.00
	1	1.00	1.00

cannot properly address the clinical demands. With a purpose of enhancing the utilization of AI in the therapeutic domain, comprehensive assistance among medical practitioners and ML analysts should be consolidated. The association will unravel the issue that ML analysts cannot acquire medical data and also to evolve more DL algorithms consistent with clinical requirements and implement them to CAD appliances, consequently enhancing the prognosis efficacy and accuracy. Therapeutic representations are distinct from other images and there are variations among diverse therapeutic images as well. These variations have an impact on the versatility of the DL model during segmentation. The noises of therapeutic images are main challenge in data pre-refining. The prevailing therapeutic image data bases are limited in scale, but the training of DL mechanism necessitates abundant data carrier that prompts the problem of over-fitting. The DL has its own deformities as it primarily considers 3 features: network organization layout, 3D data

Fig. 33. Illustrations of output as anticipated by the VGG19 model (<https://www.kaggle.com/code/enesztrk/facial-emotion-recognition-vgg19-fer2013>)Fig. 34. (a). illustration of sample pictures for both classes (b) illustrations for the test outcomes (<https://www.kaggle.com/code/lucasar/vehicle-detection-inception-xception-mobilenet>)

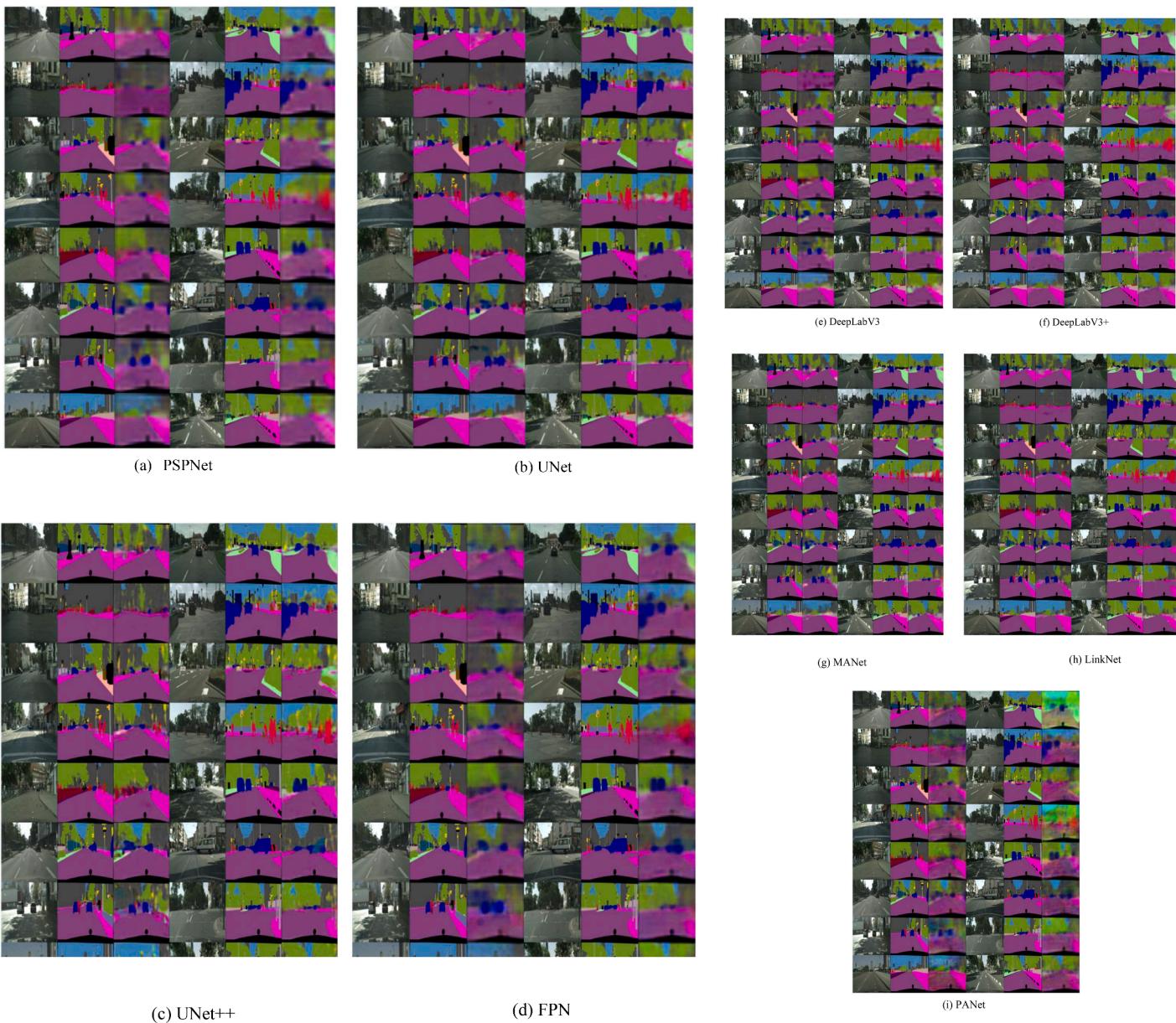
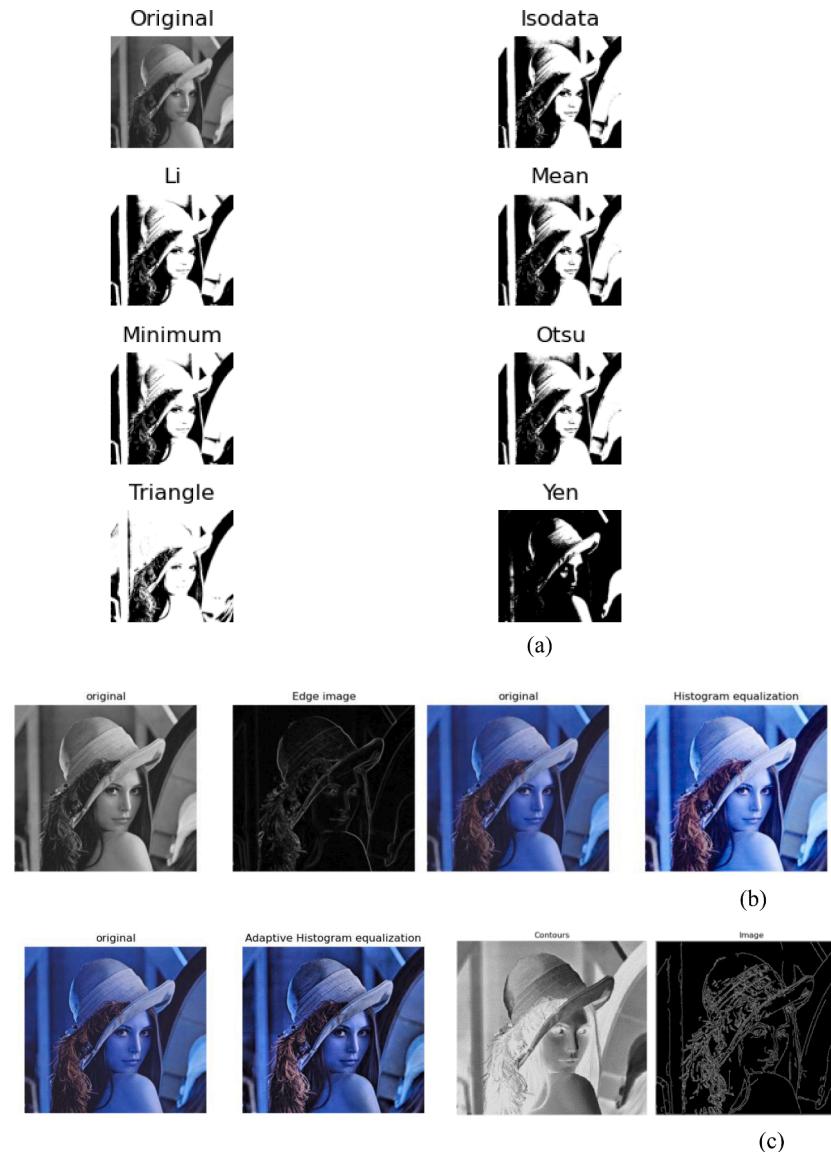


Fig. 35. Illustration of different DL based segmentation approaches (<https://github.com/divamgupta/image-segmentation-keras>)

Table 22

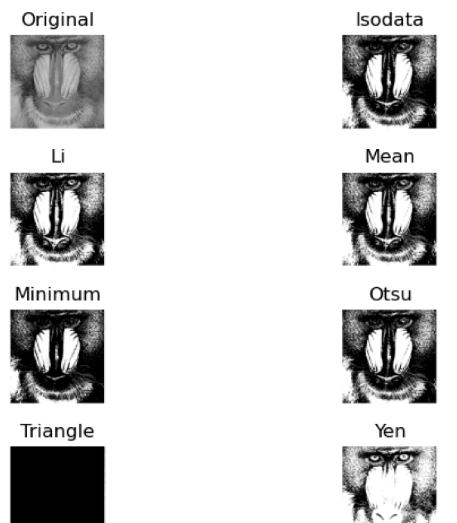
Comparative analysis of various segmentation methodologies on cityscapes database

Evaluation Metric	PSPNet	UNet	Unet++	FPN	DeeplabV3	DeeplabV3+	MANet	LinkNet	PAN
Training Loss	0.55	0.54	0.53	0.55	0.54	0.54	0.54	0.54	0.55
Training Dice	0.43	0.47	0.54	0.41	0.46	0.48	0.49	0.52	0.44
Test loss	0.57	0.56	0.56	0.57	0.57	0.56	0.56	0.56	0.57
Test Dice	0.42	0.42	0.51	0.38	0.46	0.50	0.47	0.48	0.40

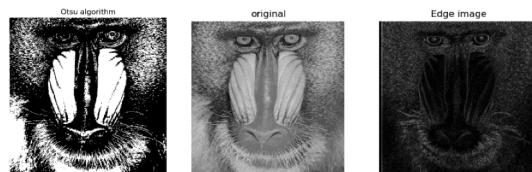
**Fig. 37.** (a). Illustration of different thresholding-based segmentation on test image-Lenna, (b)Edge based and HE segmentation (c) AHE and Contour based segmentation result on test image-Lenna

segmentation standard layout and error assessment layout. The layout of the network structure is having to be examined. The consequence of transforming the network framework is substantial and can be readily relocated to distinct functions. 3D therapeutic data can more precisely acquire the statistical details of the objective that might be disoriented when the 3D data is fragmented. So, a consideration direction is the development of 3D models for handling 3D therapeutic image data. Since the demand of 3D segmentation in therapeutic image evaluation is enhancing, there is an absolute necessity of comprehensive 3D image datasets. Weakly-supervised and unsupervised learning are growing research domains that are anticipated to be beneficial for image segmentation [339,340]. Even though DL-based methodologies have

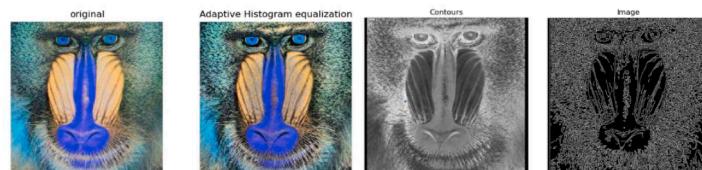
attained encouraging results on benchmarks, there exist direct questions about what actually DL representations are learning and how should one decipher the attributes learned by DL models. A beneficial research inclination that can assist in enhancing segmented picture designs is to have knowledge of a nominal neural architecture which can attain specific segmentation accuracy on a considerable dataset. Many modern segmentation models require a significant amount of memory even during the inference stage. Considerable effort has been destined for enhancing the accuracy of DL models, but as a means to fit them to certain appliances, the network needs to be uncomplicated. This can be accomplished by employing fundamental models, or by utilizing model contraction approaches, or by training a convoluted model and then



(a)



(b)



(c)

Fig. 38. (a). Illustration of different thresholding-based segmentation on test image-Mandrill, (b) Edge based and HE segmentation (c) AHE and Contour based segmentation result on test image-Mandrill.

Table 24

Comparison of PSNR values between thresholding-based segmentation techniques for test image-Lenna

Attacks	Otsu	HE	AHE	Edge	Contour	Canny
Gaussian noise	23.0788	21.4558	21.3781	21.7543	21.1547	22.1249
Salt and pepper noise	20.1205	20.1205	21.0439	20.6920	21.3020	20.1566
Speckle noise	21.6316	21.6316	22.1170	22.7182	20.6739	23.8918

utilizing knowledge refining methodologies to abridge it to memory efficient system that imitates the complicated model. DL approaches have substantially boosted segmentation accuracy owing to the competence to oversee convoluted circumstances. For acquiring such proficiency, the arrangement consistently necessitates score of marked elements to execute the training objective.

Assembling such massive dataset of analysed instances in therapeutic picture refining is frequently an enormous challenge and executing the

Table 25

Comparison of SSIM values between thresholding-based segmentation techniques for test image-Lenna

Attacks	Otsu	HE	AHE	Edge	Contour	Canny
Gaussian noise	0.6777	0.8037	0.8226	0.8055	0.7194	0.8776
Salt and pepper noise	0.6670	0.8105	0.8277	0.8168	0.7746	0.888
Speckle noise	0.6318	0.7851	0.8155	0.8024	0.6842	0.9001

Table 26

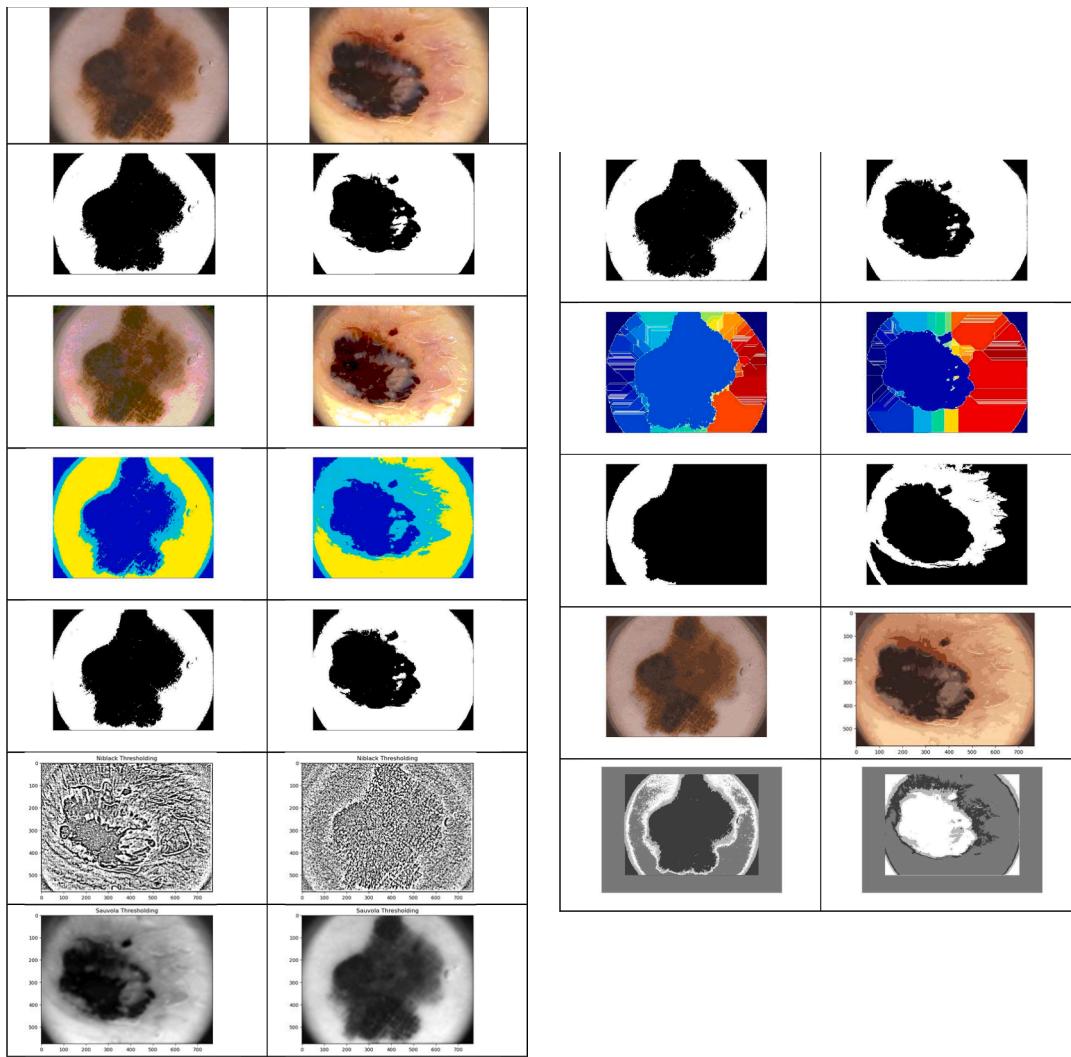
Comparison of PSNR values between thresholding-based segmentation techniques for test image-Mandrill

Attacks	Otsu	HE	AHE	Edge	Contour	Canny
Gaussian noise	22.5166	21.5147	21.4779	21.6196	21.1017	21.6303
Salt and pepper noise	20.2568	21.0064	21.0942	20.8976	21.3237	20.7680
Speckle noise	21.9294	21.2749	21.3318	22.1950	21.7381	22.8532

Table 27

Comparison of SSIM values between thresholding-based segmentation techniques for test image-Mandrill

Attacks	Otsu	HE	AHE	Edge	Contour	Canny
Gaussian noise	0.7866	0.7651	0.7722	0.7798	0.8354	0.8935
Salt and pepper noise	0.7776	0.7748	0.7810	0.7936	0.8609	0.8948
Speckle noise	0.7587	0.7381	0.7483	0.7684	0.8385	0.8932

**Fig. 36.** Illustration of different segmentation approaches on Ph2 dataset for skin cancer detection.**Table 23**

Comparative analysis of various segmentation methodologies on PH2 database

Approach	Acc	SE	SP	F1	MCC
Otsu(global)	84.2	92.5	67.5	88	63.34
Otsu (2 level)	82.5	97.5	52.5	88	60.19
Otsu (3 level)	90	96	77.5	92.7	77.15
Otsu (7 level)	84.2	92.5	67.50	88.62	63.33
Niblack	83.3	91	67.50	87.95	61.46
Sauvola	75	97.5	30.00	83.80	40.38
Watershed	81	95	55.00	87.35	57.21
FCM	81	88	67.5	86.58	57.87
K-Means	86	95	67.50	89.90	67.31
FMM	72.5	81.2	55.00	79.75	37.00

annotation on new pictures will be exhausting and extravagant. Numerous methodologies have been broadly implemented to fix this issue. For instance, the regularly practiced technique to reinforce the proportion of the training set of data is augmentation. For a non-therapeutic assessment, the efficacy of data augmentation is estimated, and the outcomes reveal that the standard modes can escalate the interpretation. Transfer learning over the effective representations enforced in the same domain is alternative to resolve the problem. Weakly supervised learning strategies [341,342,343] are applicable to manage the implications of inadequate information. Unsupervised techniques are employed to withdraw precise information and to operate this information to instruct the system that is regarded as a hybrid mode to resolve the problem. An additional problem to deal with is to assemble an appropriate group of negative specimens. For boosting the refinement power of the system on FP cases, the negative group should incorporate instances that resembles lump but not positive. It is usual in therapeutic picture refining that the evaluation of concern only engages

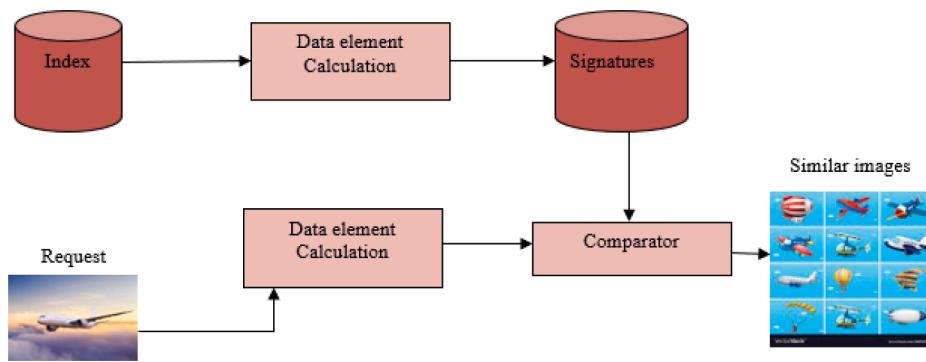


Fig 39. Illustration of content-based image retrieval [338]

a fraction of the picture. So, majority of the withdrawn quantities characterizes the background field, though the abnormalities are of increasing significance. Instructing a framework by this data induces the trained system being influenced of the background and confined to local minima. The interpretation for this problem is rebalancing the specimen. Mechanized transformation of rebalanced specimen has been fabricated by utilizing DI [344,345]. Nevertheless, the efficacy is confined with reference to immoderate class variation [346]. Above all, through the conception of the training set, a compliance channel can be fixed possessing equitable patches off the background and foreground [347]. Distinct methodology to cover the problem is sampled loss where the loss is computed merely for specified regions and not the full picture.

CRediT authorship contribution statement

Khushmeen Kaur Brar: Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Data curation, Conceptualization. **Bhawna Goyal:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Conceptualization. **Ayush Dogra:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology. **Mohammed Ahmed Mustafa:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Rana Majumdar:** Writing – review & editing, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Ahmed Alkhayyat:** Writing – review & editing, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Vinay Kukreja:** Visualization, Validation, Supervision, Resources, Conceptualization.

Declaration of competing interest

There is no conflict of interest amongst authors

Data availability

No data was used for the research described in the article.

References

- [1] B. Jin, L. Cruz, N. Goncalves, Deep facial diagnosis: deep transfer learning from face recognition to facial diagnosis, *IEEE Access*. 8 (2020) 123649–123661, <https://doi.org/10.1109/ACCESS.2020.3005687>.
- [2] X.P. Burgos-Artizru, A. Ribeiro, A. Tellaeche, G. Pajares, C. Fernández-Quintanilla, Analysis of natural images processing for the extraction of agricultural elements, *Image Vis. Comput.* 28 (1) (2010) 138–149, <https://doi.org/10.1016/j.imavis.2009.05.009>.
- [3] F. Pedreschi, J. León, D. Mery, P. Moyano, Development of a computer vision system to measure the color of potato chips, *Food Res. Int.* 39 (10) (2006) 1092–1098, <https://doi.org/10.1016/j.foodres.2006.03.009>.
- [4] D. Unay, B. Gosselin, Thresholding-based segmentation and apple grading by machine vision, in: *13th Eur. Signal Process. Conf. EUSIPCO 2005*, 2005, pp. 926–929.
- [5] H.P. Ng, S.H. Ong, K.W.C. Foong, P.S. Goh, W.L. Nowinski, Medical image segmentation using k-means clustering and improved watershed algorithm, *Proc. IEEE Southwest Symp. Image Anal. Interpret.* 2006 (2006) 61–65, <https://doi.org/10.1109/ssai.2006.1633722>.
- [6] M. Y. Xu Guang-zhu, ZHANG Zai-feng, “An Image Segmentation based method for iris feature extraction,” vol. 15, no. 1, 2008.
- [7] P. Samant, R. Agarwal, Machine learning techniques for medical diagnosis of diabetes using iris images, *Comput. Methods Programs Biomed.* 157 (2018) 121–128, <https://doi.org/10.1016/j.cmpb.2018.01.004>.
- [8] M. Urschler, A. Bornik, E. Scheurer, K. Yen, H. Bischof, D. Schmalstieg, Forensic-case analysis: From 3D imaging to interactive visualization, *IEEE Comput. Graph. Appl.* 32 (4) (2012) 79–87, <https://doi.org/10.1109/MCG.2012.75>.
- [9] R. Wanat, A problem of automatic segmentation of digital dental panoramic x-ray images for forensic human identification, *Cent. Eur.* (2011).
- [10] E.R. DAVIES, *Computer and Machine Vision: Theory, Algorithms, Practicalities* (1) (2012).
- [11] R.B. Milan Sonka, Vaclav Hlavac, *Image Processing, Analysis, and Machine Vision* (2013).
- [12] R.C. Gonzalez, R.E. Woods, *Digital image processing* (2018).
- [13] S. Arora, J. Acharya, A. Verma, P.K. Panigrahi, Multilevel thresholding for image segmentation through a fast statistical recursive algorithm, *Pattern Recognit. Lett.* 29 (2) (2008) 119–125, <https://doi.org/10.1016/j.patrec.2007.09.005>.
- [14] P.K. Sahoo, S. Soltani, A.K.C. Wong, A survey of thresholding techniques, *Comput. Vision, Graph. Image Process.* 41 (2) (1988) 233–260, [https://doi.org/10.1016/0734-189X\(88\)90022-9](https://doi.org/10.1016/0734-189X(88)90022-9).
- [15] S.-M. G., P.D.T.C.-I Chang, Y. Du, J. Wang, Survey and comparative analysis of entropy and relative entropy thresholding techniques, *Image* (Rochester, N.Y.) (November 1995) (2014) 205–212, <https://doi.org/10.1049/ip-vis>.
- [16] T. Kurita, N. Otsu, N. Abdelmalek, Maximum likelihood thresholding based on, *Pattern. Recognit.* 25 (10) (1992) 1231–1240 [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/003132039290024D>.
- [17] J.N. Kapur, P.K. Sahoo, A.K.C. Wong, A new method for gray-level picture thresholding using the entropy of the histogram, *Comput. Vision, Graph. Image Process.* 29 (3) (1985) 273–285, [https://doi.org/10.1016/0734-189X\(85\)90125-2](https://doi.org/10.1016/0734-189X(85)90125-2).
- [18] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Trans. Syst. Man Cybern.* 9 (1) (1979) 62–66.
- [19] P. Ghamisi, M.S. Couceiro, F.M.L. Martins, J.A. Benediktsson, Multilevel image segmentation based on fractional-order darwinian particle swarm optimization, *IEEE Trans. Geosci. Remote Sens.* 52 (5) (2014) 2382–2394, <https://doi.org/10.1109/TGRS.2013.2260552>.
- [20] A. Dirami, K. Hammouche, M. Diaf, P. Siarry, Fast multilevel thresholding for image segmentation through a multiphase level set method, *Signal Process.* 93 (1) (2013) 139–153, <https://doi.org/10.1016/j.sigpro.2012.07.010>.
- [21] B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation, *J. Electron. Imaging* 13 (1) (2004) 146, <https://doi.org/10.1117/1.1631315>.
- [22] B. Akay, A study on particle swarm optimization and artificial bee colony algorithms for multilevel thresholding, *Appl. Soft Comput.* J. 13 (6) (2013) 3066–3091, <https://doi.org/10.1016/j.asoc.2012.03.072>.
- [23] K. Hamrnouche, M. Diaf, P. Siarry, A comparative study of various meta-heuristic techniques applied to the multilevel thresholding problem, *Eng. Appl. Artif. Intell.* 23 (5) (2010) 676–688, <https://doi.org/10.1016/j.engappai.2009.09.011>.
- [24] J. Zhang, H. Li, Z. Tang, Q. Lu, X. Zheng, J. Zhou, An improved quantum-inspired genetic algorithm for image multilevel thresholding segmentation, *Math. Probl. Eng.* 2014 (2014), <https://doi.org/10.1155/2014/295402>.
- [25] K. Hammouche, M. Diaf, and P. Siarry, “A multilevel automatic thresholding method based on a genetic algorithm for a fast image segmentation,” vol. 109, pp. 163–175, 2008, [10.1016/j.cviu.2007.09.001](https://doi.org/10.1016/j.cviu.2007.09.001).
- [26] K. Tang, X. Yuan, T. Sun, J. Yang, S. Gao, An improved scheme for minimum cross entropy threshold selection based on genetic algorithm, *Knowledge-Based Syst* 24 (8) (2011) 1131–1138, <https://doi.org/10.1016/j.knosys.2011.02.013>.

- [27] R. by, K.P. Storn, Differential Evolution - A simple and efficient adaptive scheme for global optimization over continuous spaces, *Entomol. Exp. Appl.* 103 (3) (2002) 239–248, <https://doi.org/10.1023/A>.
- [28] E. Cuevas, D. Zaldivar, M. Pérez-Cisneros, A novel multi-threshold segmentation approach based on differential evolution optimization, *Expert Syst. Appl.* 37 (7) (2010) 5265–5271, <https://doi.org/10.1016/j.eswa.2010.01.013>.
- [29] M. Ali, P. Siarry, M. Pant, An efficient Differential Evolution based algorithm for solving multi-objective optimization problems, *Eur. J. Oper. Res.* 217 (2) (2012) 404–416, <https://doi.org/10.1016/j.ejor.2011.09.025>.
- [30] W. Tao, H. Jin, L. Liu, Object segmentation using ant colony optimization algorithm and fuzzy entropy, *Pattern Recognit. Lett.* 28 (7) (2007) 788–796, <https://doi.org/10.1016/j.patrec.2006.11.007>.
- [31] P.D. Sathy, R. Kayalvizhi, Optimal multilevel thresholding using bacterial foraging algorithm, *Expert Syst. Appl.* 38 (12) (2011) 15549–15564, <https://doi.org/10.1016/j.eswa.2011.06.004>.
- [32] N. Sanyal, A. Chatterjee, S. Munshi, An adaptive bacterial foraging algorithm for fuzzy entropy based image segmentation, *Expert Syst. Appl.* 38 (12) (2011) 15489–15498, <https://doi.org/10.1016/j.eswa.2011.06.011>.
- [33] M.A. Bakhtshali, M. Shamsi, Segmentation of color lip images by optimal thresholding using bacterial foraging optimization (BFO), *J. Comput. Sci.* 5 (2) (2014) 251–257, <https://doi.org/10.1016/j.jocs.2013.07.001>.
- [34] D. Oliva, E. Cuevas, G. Pajares, D. Zaldivar, M. Perez-Cisneros, Multilevel thresholding segmentation based on harmony search optimization, *J. Appl. Math.* 2013 (2013), <https://doi.org/10.1155/2013/575414>.
- [35] V. O. c Diego Oliva a, Erik Cuevash, Gonzalo Pajares a, Daniel Zaldivar b, “A multilevel thresholding algorithm using Electro-magnetism Optimization,” pp. 357–381, 2014.
- [36] M.H. Horng, A multilevel image thresholding using the honey bee mating optimization, *Appl. Math. Comput.* 215 (9) (2010) 3302–3310, <https://doi.org/10.1016/j.amc.2009.10.018>.
- [37] M.H. Horng, Multilevel minimum cross entropy threshold selection based on the honey bee mating optimization, *Expert Syst. Appl.* 37 (6) (2010) 4580–4592, <https://doi.org/10.1016/j.eswa.2009.12.050>.
- [38] M.H. Horng, R.J. Liou, Multilevel minimum cross entropy threshold selection based on the firefly algorithm, *Expert Syst. Appl.* 38 (12) (2011) 14805–14811, <https://doi.org/10.1016/j.eswa.2011.05.069>.
- [39] X.S. Yang, Firefly algorithms for multimodal optimization, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 5792 LNCS (March 2010) (2009) 169–178, https://doi.org/10.1007/978-3-642-04944-6_14.
- [40] X.-S.S. Yang, M. Karamanoglu, Nature-Inspired Metaheuristic Algorithms Second Edition 4 (C) (2013) [Online]. Available: <http://www.sciencedirect.com/science/article/pii/B9780124051638000016%0Ahttp://linkinghub.elsevier.com/retrieve/pii/B978012405163800002X>.
- [41] D. KARABOGA, “An idea based on honey bee swarm for numerical optimization,” no. March, pp. 25–27, 2005.
- [42] A.K. Bhandari, V.K. Singh, A. Kumar, G.K. Singh, Cuckoo search algorithm and wind driven optimization based study of satellite image segmentation for multilevel thresholding using Kapur’s entropy, *Expert Syst. Appl.* 41 (7) (2014) 3538–3560, <https://doi.org/10.1016/j.eswa.2013.10.059>.
- [43] A.K. Bhandari, V. Soni, A. Kumar, G.K. Singh, Artificial Bee Colony-based satellite image contrast and brightness enhancement technique using DWT-SVD, *Int. J. Remote Sens.* 35 (5) (2014) 1601–1624, <https://doi.org/10.1080/01431161.2013.876518>.
- [44] V. Soni, A.K. Bhandari, A. Kumar, G.K. Singh, Improved sub-band adaptive thresholding function for denoising of satellite image based on evolutionary algorithms, *IET Signal Process* 7 (8) (2013) 720–730, <https://doi.org/10.1049/iet-spr.2013.0139>.
- [45] E. Cuevas, V. Osuna-Enciso, D. Zaldivar, M. Pérez-Cisneros, H. Sossa, Multithreshold segmentation based on artificial immune systems, *Math. Probl. Eng.* 2012 (2012), <https://doi.org/10.1155/2012/874761>.
- [46] E. Cuevas, F. Sención, D. Zaldivar, M. Pérez-Cisneros, H. Sossa, A multi-threshold segmentation approach based on artificial bee colony optimization, *Appl. Intell.* 37 (3) (2012) 321–336, <https://doi.org/10.1007/s10489-011-0330-z>.
- [47] M. Ma, J. Liang, M. Guo, Y. Fan, Y. Yin, SAR image segmentation based on artificial bee colony algorithm, *Appl. Soft Comput.* 11 (8) (2011) 5205–5214, <https://doi.org/10.1016/j.asoc.2011.05.039>.
- [48] S. Kumar, P. Kumar, T.K. Sharma, M. Pant, Bi-level thresholding using PSO, Artificial Bee Colony and MRLDE embedded with Otsu method, *Memetic Comput* 5 (4) (2013) 323–334, <https://doi.org/10.1007/s12293-013-0123-5>.
- [49] Y.Bin Zhu, Overview of particle swarm optimization, *Appl. Mech. Mater.* 543–547 (2014) 1597–1600, <https://doi.org/10.4028/www.scientific.net/AMM.543-547.1597>.
- [50] M. Maitra, A. Chatterjee, A hybrid cooperative-comprehensive learning based PSO algorithm for image segmentation using multilevel thresholding, *Expert Syst. Appl.* 34 (2) (2008) 1341–1350, <https://doi.org/10.1016/j.eswa.2007.01.002>.
- [51] P.Y. Yin, Multilevel minimum cross entropy threshold selection based on particle swarm optimization, *Appl. Math. Comput.* 184 (2) (2007) 503–513, <https://doi.org/10.1016/j.amc.2006.06.057>.
- [52] H. Gao, W. Xu, J. Sun, Y. Tang, Multilevel thresholding for image segmentation through an improved quantum-behaved particle swarm algorithm, *IEEE Trans. Instrum. Meas.* 59 (4) (2010) 934–946, <https://doi.org/10.1109/TIM.2009.2030931>.
- [53] Z. Bayraktar, J.P. Turpin, D.H. Werner, Nature-inspired optimization of high-impedance metasurfaces with ultrasmall interwoven unit cells, *IEEE Antennas Wirel. Propag. Lett.* 10 (2011) 1563–1566, <https://doi.org/10.1109/LAWP.2011.2178224>.
- [54] Z. Bayraktar, M. Komurcu, J.A. Bossard, D.H. Werner, The wind driven optimization technique and its application in electromagnetics, *IEEE Trans. Antennas Propag.* 61 (5) (2013) 2745–2757, <https://doi.org/10.1109/TAP.2013.2238654>.
- [55] J. Kittler, J. Illingworth, Minimum error thresholding, *Pattern. Recognit.* 19 (1) (1986) 41–47, [https://doi.org/10.1016/0031-3203\(86\)90030-0](https://doi.org/10.1016/0031-3203(86)90030-0).
- [56] E. Zahara, S.K.S. Fan, D.M. Tsai, Optimal multi-thresholding using a hybrid optimization approach, *Pattern Recognit. Lett.* 26 (8) (2005) 1082–1095, <https://doi.org/10.1016/j.patrec.2004.10.003>.
- [57] S.K.S. Fan, Y. Lin, A multi-level thresholding approach using a hybrid optimal estimation algorithm, *Pattern Recognit. Lett.* 28 (5) (2007) 662–669, <https://doi.org/10.1016/j.patrec.2006.11.005>.
- [58] T. Kurban, P. Civicioglu, R. Kurban, E. Besdok, Comparison of evolutionary and swarm based computational techniques for multilevel color image thresholding, *Appl. Soft Comput.* 13 (2014) 128–143, <https://doi.org/10.1016/j.asoc.2014.05.037>.
- [59] A.K. Bhandari, A. Kumar, G.K. Singh, Modified artificial bee colony based computationally efficient multilevel thresholding for satellite image segmentation using Kapur’s, Otsu and Tsallis functions, *Expert Syst. Appl.* 42 (3) (2015) 1573–1601, <https://doi.org/10.1016/j.eswa.2014.09.049>.
- [60] A.K. Bhandari, A. Kumar, G.K. Singh, Tsallis entropy based multilevel thresholding for colored satellite image segmentation using evolutionary algorithms, *Expert Syst. Appl.* 42 (22) (2015) 8707–8730, <https://doi.org/10.1016/j.eswa.2015.07.025>.
- [61] H. Peng, F. Long, Z. Chi, W. Su, A hierarchical distributed genetic algorithm for image segmentation, *Proc. 2000 Congr. Evol. Comput. CEC 2000* 1 (2000) 272–276, <https://doi.org/10.1109/CEC.2000.870306>.
- [62] G.Lo Bosco, A genetic algorithm for image segmentation, in: *Proc. - 11th Int. Conf. Image Anal. Process. ICIAP 2001*, 2001, pp. 262–266, <https://doi.org/10.1109/ICIAP.2001.957019>.
- [63] W.B. Tao, J.W. Tian, J. Liu, Image segmentation by three-level thresholding based on maximum fuzzy entropy and genetic algorithm, *Pattern Recognit. Lett.* 24 (16) (2003) 3069–3078, [https://doi.org/10.1016/S0167-8655\(03\)00166-1](https://doi.org/10.1016/S0167-8655(03)00166-1).
- [64] L. Cao, P. Bao, Z. Shi, The strongest schema learning GA and its application to multilevel thresholding, *Image Vis. Comput.* 26 (5) (2008) 716–724, <https://doi.org/10.1016/j.imavis.2007.08.007>.
- [65] S. Manikandan, K. Ramar, M. Willjuice Iruthayaran, K.G. Srinivasagan, Multilevel thresholding for segmentation of medical brain images using real coded genetic algorithm, *Meas. J. Int. Meas. Confed.* 47 (1) (2014) 558–568, <https://doi.org/10.1016/j.measurement.2013.09.031>.
- [66] B. S. S., V.K.P. Ganesan, V. Rajini, Unsupervised segmentation of satellite images based on neural network and genetic algorithm, *Adv. Intell. Syst. Comput.* 309 AISC (VOLUME 2) (2015), <https://doi.org/10.1007/978-81-322-2009-1>.
- [67] E. Bonabeau, Swarm intelligence: from natural to artificial systems, *Complex. 5 Quest. (August 2014)* (2008) 21–26.
- [68] S.P. Duraisamy, R. Kayalvizhi, A new multilevel thresholding method using swarm intelligence algorithm for image segmentation, *J. Intell. Learn. Syst. Appl.* 02 (03) (2010) 126–138, <https://doi.org/10.4236/jilsa.2010.23016>.
- [69] H. Gao, C.M. Pun, S. Kwong, An efficient image segmentation method based on a hybrid particle swarm algorithm with learning strategy, *Inf. Sci. (Ny).* 369 (2016) 500–521, <https://doi.org/10.1016/j.ins.2016.07.017>.
- [70] A.H. Elsheikh, M.Abd Elaziz, Review on applications of particle swarm optimization in solar energy systems, *Int. J. Environ. Sci. Technol.* 16 (2) (2019) 1159–1170, <https://doi.org/10.1007/s13762-018-1970-x>.
- [71] Y. Zhang, L. Wu, Optimal multi-level thresholding based on maximum Tsallis entropy via an artificial bee colony approach, *Entropy* 13 (4) (2011) 841–859, <https://doi.org/10.3390/e13040841>.
- [72] M.H. Horng, Multilevel thresholding selection based on the artificial bee colony algorithm for image segmentation, *Expert Syst. Appl.* 38 (11) (2011) 13785–13791, <https://doi.org/10.1016/j.eswa.2011.04.180>.
- [73] M.-H. Horng, Multilevel minimum cross entropy image thresholding using artificial bee colony algorithm, *Telkomnika Indones. J. Electr. Eng.* 11 (9) (2013), <https://doi.org/10.11591/telkomnika.v11i9.3273>.
- [74] M.A. Awadallah, M.A. Al-Betar, A.L. Bolaji, E.M. Alsukhni, H. Al-Zoubi, Natural selection methods for artificial bee colony with new versions of onlooker bee, *Soft. comput.* 23 (15) (2019) 6455–6494, <https://doi.org/10.1007/s00500-018-3299-2>.
- [75] P. D. Sathy and R. Kayalvizhi, “Optimum Multilevel Image Thresholding Based on Tsallis Entropy Method with Bacterial Foraging Algorithm,” vol. 7, no. 5, pp. 336–343, 2010.
- [76] P.D. Sathy, R. Kayalvizhi, Amended bacterial foraging algorithm for multilevel thresholding of magnetic resonance brain images, *Meas. J. Int. Meas. Confed.* 44 (10) (2011) 1828–1848, <https://doi.org/10.1016/j.measurement.2011.09.005>.
- [77] P.D. Sathy, R. Kayalvizhi, Modified bacterial foraging algorithm based multilevel thresholding for image segmentation, *Eng. Appl. Artif. Intell.* 24 (4) (2011) 595–615, <https://doi.org/10.1016/j.engappai.2010.12.001>.
- [78] K. Tang, X. Xiao, J. Wu, J. Yang, L. Luo, An improved multilevel thresholding approach based modified bacterial foraging optimization, *Appl. Intell.* 46 (1) (2017) 214–226, <https://doi.org/10.1007/s10489-016-0832-9>.
- [79] S. Sarkar, S. Das, Multilevel image thresholding based on 2D histogram and maximum tsallis entropy - A differential evolution approach, *IEEE Trans. Image Process.* 22 (12) (2013) 4788–4797, <https://doi.org/10.1109/TIP.2013.2277832>.

- [80] K. Charansiriphaisan, S. Chiewchanwattana, K. Sunat, A global multilevel thresholding using differential evolution approach, *Math. Probl. Eng.* 2014 (2014), <https://doi.org/10.1155/2014/974024>.
- [81] W. Cao, J. Li, J. Liu, P. Zhang, Two improved segmentation algorithms for whole cardiac CT sequence images, in: *Proc. - 2016 9th Int. Congr. Image Signal Process. Biomed. Eng. Informatics, CISPB-MEI 2016*, 2017, pp. 346–351, <https://doi.org/10.1109/CISPB-MEI.2016.7852734>.
- [82] F.Y. Shih, S. Cheng, Adaptive mathematical morphology for edge linking, *Inf. Sci. (Ny)*. 167 (1–4) (2004) 9–21, <https://doi.org/10.1016/j.ins.2003.07.020>.
- [83] D.R. Martin, C.C. Fowlkes, J. Malik, Learning to Detect Natural Image Boundaries Using Brightness and Texture, in: *NIPS 2002 Proc. 15th Int. Conf. Neural Inf. Process. Syst.* 26, 2002, pp. 1255–1262.
- [84] G. Castilla, G.J. Hay, Image objects and geographic objects, *Lect. Notes Geoinf. Cartogr.* 0 (9783540770572) (2008) 91–110, https://doi.org/10.1007/978-3-540-77058-9_5.
- [85] Z. Chen, Z. Zhao, P. Gong, B. Zeng, A new process for the segmentation of high resolution remote sensing imagery, *Int. J. Remote Sens.* 27 (22) (2006) 4991–5001, <https://doi.org/10.1080/01431160600658131>.
- [86] K. Sahin, L. Ulusoy, Automatic Multi-Scale Segmentation Of High Spatial Resolution Satellite Images Using Watersheds Metuvision Lab., Dept. of Electrical and Electronics Eng., METU, Ankara, Turkey Dept. of Receiver and Transmitter Technologies, REWIS Division, ASELSAN, in: *2013 IEEE Int. Geosci. Remote Sens. Symp. IGARSS*, 2013, pp. 2505–2508.
- [87] C.H. Fosgate, H. Krim, W.W. Irving, W.C. Karl, A.S. Willsky, Multiscale segmentation and anomaly enhancement of SAR imagery, *IEEE Trans. Image Process.* 6 (1) (1997) 7–20, <https://doi.org/10.1109/83.552077>.
- [88] P. Andrey, P. Tarroux, Unsupervised segmentation of markov random field modeled textured images using selectionist relaxation, *IEEE Trans. Pattern Anal. Mach. Intell.* 20 (3) (1998) 252–262, <https://doi.org/10.1109/34.667883>.
- [89] J. CANNY, A computational approach to edge detection, *Readings Comput. Vis.* (1987) 184–203, <https://doi.org/10.1016/b978-0-08-051581-6.50024-6>.
- [90] R. Đurković, K. Kaneda, H. Yamashita, Dynamic contour: A texture approach and contour operations, *Vis. Comput.* 11 (6) (1995) 277–289, <https://doi.org/10.1007/BF01898405>.
- [91] H. Farid, E.P. Simoncelli, Optimally rotation-equivariant directional derivative kernels, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 1296 (1997) 207–214, https://doi.org/10.1007/3-540-63460-6_119.
- [92] A. W., D.T. Michael Kass, Snakes: active contour models, *Int. J. Comput. Vis.* (1988) 321–331, <https://doi.org/10.1016/j.procs.2018.10.231>.
- [93] F. Leymarie, M.D. Levine, Tracking deformable objects in the plane using an active contour model, *IEEE Trans. Pattern Anal. Mach. Intell.* 15 (6) (1993) 617–634, <https://doi.org/10.1109/34.216733>.
- [94] D. Marr, E. Hildreth, Theory of edge detection, *Proc. R. Soc. London - Biol. Sci.* 207 (1167) (1980) 187–217, <https://doi.org/10.1098/rspb.1980.0020>.
- [95] S.D. Mayunga, D.J. Coleman, Y. Zhang, A semi-automated approach for extracting buildings from QuickBird imagery applied to informal settlement mapping, *Int. J. Remote Sens.* 28 (10) (2007) 2343–2357, <https://doi.org/10.1080/01431160600868474>.
- [96] P. Meer, B. Georgescu, Edge detection with embedded confidence, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (12) (2001) 1351–1365, <https://doi.org/10.1109/34.977560>.
- [97] J. Peng, D. Zhang, Y. Liu, An improved snake model for building detection from urban aerial images, *Pattern Recognit. Lett.* 26 (5) (2005) 587–595, <https://doi.org/10.1016/j.patrec.2004.09.033>.
- [98] G.S. Robinson, Edge detection by compass gradient masks, *Comput. Graph. Image Process* 6 (5) (1977) 492–501, [https://doi.org/10.1016/S0146-664X\(77\)80024-5](https://doi.org/10.1016/S0146-664X(77)80024-5).
- [99] M.M. Trivedi, J.C. Bezdek, Low-level segmentation of aerial images with fuzzy clustering, *IEEE Trans. Syst. Man Cybern.* 16 (4) (1986) 589–598, <https://doi.org/10.1109/TSMC.1986.289264>.
- [100] T. S. and R. C. B.S. Manjunath, “Stochastic and Deterministic Networks for Texture Segmentation.”.
- [101] L. Lucchesi, S.K. Mitray, Color image segmentation: A state-of-the-art survey, in: *Proc. Indian Natl. Sci. Acad. (INSA-A)*, Delhi, Indian Natl Sci Acad 67, 2001, pp. 207–221 [Online]. Available, <http://ultra.sdlk.free.fr/docs/Image-Processing/filters/Color-Image-Segmentation-A-State-of-the-Art-Survey.pdf%5Cnhttp://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.84.4896>.
- [102] J.A. Jiang, C.L. Chuang, Y.L. Lu, C.S. Fahn, Mathematical-morphology-based edge detectors for detection of thin edges in low-contrast regions, *IET. Image Process.* 1 (3) (2007) 269–277, <https://doi.org/10.1049/iet-ipr:20060273>.
- [103] D.S. Lu, C.C. Chen, Edge detection improvement by ant colony optimization, *Pattern Recognit. Lett.* 29 (4) (2008) 416–425, <https://doi.org/10.1016/j.patrec.2007.10.021>.
- [104] D.H. Ballard, Generalizing the hough transform to detect arbitrary shapes, *Pattern. Recognit.* 13 (2) (1981) 111–122, <https://doi.org/10.23919/CC55666.2022.9902668>.
- [105] O. G. and P. F. Whelan, “A Computational Approach for Edge Linking”.
- [106] L. V., P. Soille, Watersheds in digital spaces: An efficient algorithm based on immersion simulations, *IEEE Trans. Pattern. Anal. Mach. Intell.* 13 (3) (1991) 583–598, <https://doi.org/10.1080/00207450390162137>.
- [107] B. Kaur, A. Garg, Mathematical morphological edge detection for remote sensing images, in: *ICECT 2011 - 2011 3rd Int. Conf. Electron. Comput. Technol.* 5, 2011, pp. 324–327, <https://doi.org/10.1109/ICECTECH.2011.5942012>.
- [108] F. Meyer, S. Beucher, Morphological segmentation, *J. Vis. Commun. Image Represent.* 1 (1) (1990) 21–46, [https://doi.org/10.1016/1047-3203\(90\)90014-M](https://doi.org/10.1016/1047-3203(90)90014-M).
- [109] X. Munoz, J. Freixenet, X. Cuff, J. Martí, Strategies for image segmentation combining region and boundary information, *Pattern Recognit. Lett.* 24 (1–3) (2003) 375–392, [https://doi.org/10.1016/S0167-8655\(02\)00262-3](https://doi.org/10.1016/S0167-8655(02)00262-3).
- [110] V. Mezaris, I. Kompatiari, M.G. Strintzis, Still image segmentation tools for object-based multimedia applications, *Int. J. Pattern Recognit. Artif. Intell.* 18 (4) (2004) 701–725, <https://doi.org/10.1142/S0218001404003393>.
- [111] A.P. Carleer, O. Debeir, E. Wolff, Assessment of very high spatial resolution satellite image segmentations, *Photogramm. Eng. Remote Sensing* 71 (11) (2005) 1285–1294, <https://doi.org/10.14358/PERS.71.11.1285>.
- [112] J. Yang, Y. He, J. Caspersen, Region merging using local spectral angle thresholds: A more accurate method for hybrid segmentation of remote sensing images, *Remote Sens. Environ.* 190 (2017) 137–148, <https://doi.org/10.1016/j.rse.2016.12.011>.
- [113] G. Meinel, M. Neubert, *A Comparison of Segmentation Programs for High Resolution Remote Sensing Data*, in: *XXth ISPRS Congr. Tech. Comm. IV*, 2004, p. 6.
- [114] T. Zuva, O.O. Olugbara, S.O. Ojo, S.M. Ngwira, Image segmentation, available techniques, developments and open issues, *Can. J. Image Process. Comput. Vis.* 2 (3) (2011) 20–29 [Online]. Available: https://www.researchgate.net/profile/Tomas_Zuva/publication/264854010_Image_Segmentation_Available_Techniques_Developments_and_Open_Issues/links/563b097708ae0d0531dcc7a2/Image-Segmentation-Available-Techniques-Developments-and-Open-Issues.pdf.
- [115] J. Weickert, Efficient image segmentation using partial differential equations and morphology, *Patter. Recognit.* 34 (9) (2001) 1813–1824, [https://doi.org/10.1016/S0031-3203\(00\)00109-6](https://doi.org/10.1016/S0031-3203(00)00109-6).
- [116] T. Zheng, Z. Duan, J. Wang, G. Lu, S. Li, Z. Yu, Research on distance transform and neural network lidar information sampling classification-based semantic segmentation of 2D indoor room maps, *Sensors (Switzerland)* 21 (4) (2021) 1–20, <https://doi.org/10.3390/s21041365>.
- [117] Y.I. Ohta, T. Kanade, T. Sakai, Color information for region segmentation, *Comput. Graph. Image Process.* 13 (3) (1980) 222–241, [https://doi.org/10.1016/0146-664X\(80\)90047-7](https://doi.org/10.1016/0146-664X(80)90047-7).
- [118] T. Kavzoglu, H. Tonbul, A comparative study of segmentation quality for multi-resolution segmentation and watershed transform, in: *Proc. 8th Int. Conf. Recent Adv. Sp. Technol. RAST 2017*, 2017, pp. 113–117, <https://doi.org/10.1109/RAST.2017.8002984>.
- [119] A. Tremeau, N. Borel, A region growing and merging algorithm to color segmentation, *Pattern. Recognit.* 30 (7) (1997) 1191–1203, [https://doi.org/10.1016/S0031-3203\(96\)00147-1](https://doi.org/10.1016/S0031-3203(96)00147-1).
- [120] J. Fan, D.K.Y. Yau, A.K. Elmagarmid, W.G. Aref, Automatic image segmentation by integrating color-edge extraction and seeded region growing, *IEEE Trans. Image Process.* 10 (10) (2001) 1454–1466, <https://doi.org/10.1109/83.951532>.
- [121] R. Adams, L. Bischof, Seeded region growing, *IEEE Trans. Pattern Anal. Mach. Intell.* 16 (6) (1994) 641–647, <https://doi.org/10.1109/34.295913>.
- [122] A. Mehner, P. Jackway, An improved seeded region growing algorithm, *Pattern Recognit. Lett.* 18 (10) (1997) 1065–1071, [https://doi.org/10.1016/S0167-8655\(97\)00131-1](https://doi.org/10.1016/S0167-8655(97)00131-1).
- [123] F.Y. Shih, S. Cheng, Automatic seeded region growing for color image segmentation, *Image Vis. Comput.* 23 (10) (2005) 877–886, <https://doi.org/10.1016/j.imavis.2005.05.015>.
- [124] O.P. Verma, M. Hanmandlu, S. Susan, M. Kulkarni, P.K. Jain, A simple single seeded region growing algorithm for color image segmentation using adaptive thresholding, in: *Proc. - 2011 Int. Conf. Commun. Syst. Netw. Technol. CSNT 2011*, 2011, pp. 500–503, <https://doi.org/10.1109/CSNT.2011.107>.
- [125] X. Zhang, P. Xiao, X. Feng, J. Wang, Z. Wang, Hybrid region merging method for segmentation of high-resolution remote sensing images, *ISPRS J. Photogramm. Remote Sens.* 98 (2014) 19–28, <https://doi.org/10.1016/j.isprsjprs.2014.09.011>.
- [126] Y. Byun, D. Kim, J. Lee, Y. Kim, A framework for the segmentation of high-resolution satellite imagery using modified seeded-region growing and region merging, *Int. J. Remote Sens.* 32 (16) (2011) 4589–4609, <https://doi.org/10.1080/01431161.2010.489066>.
- [127] R. Nock, F. Nielsen, Statistical region merging, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (11) (2004) 1452–1458, <https://doi.org/10.1109/TPAMI.2004.110>.
- [128] Y. Xiaohan, J. Ylä-Jääski, O. Huttunen, T. Vehkomäki, O. Sipilä, T. Katila, Image segmentation combining region growing and edge detection, in: *Proc. - Int. Conf. Pattern Recognit.* 3, 1992, pp. 481–484, <https://doi.org/10.1109/ICPR.1992.202029>.
- [129] G.J. Hay, G. Castilla, Geographic object-based image analysis (GEOBIA): A new name for a new discipline, *Lect. Notes Geoinf. Cartogr.* 0 (9783540770572) (2008) 75–89, https://doi.org/10.1007/978-3-540-77058-9_4.
- [130] M. BAATZ, “Mafiresolution Segmentation:an optimization approach for high quality multi-scale image segmentation,” vol. 13, no. 4, pp. 1–12, 2000, [10.1207/s15326888chc1304_3](https://doi.org/10.1207/s15326888chc1304_3).
- [131] H. Gao, Y. Tang, L. Jing, H. Li, H. Ding, A novel unsupervised segmentation quality evaluation method for remote sensing images, *Sensors (Switzerland)* 17 (10) (2017) 1–22, <https://doi.org/10.3390/s17102427>.
- [132] B. Kalantar, S. Bin Mansor, M.I. Sameen, B. Pradhan, H.Z.M. Shafri, Drone-based land-cover mapping using a fuzzy unordered rule induction algorithm integrated into object-based image analysis, *Int. J. Remote Sens.* 38 (8–10) (2017) 2535–2556, <https://doi.org/10.1080/01431161.2016.1277043>.
- [133] G. Mallinis, N. Koutsias, M. Tsakiri-Strati, M. Karteris, Object-based classification using Quickbird imagery for delineating forest vegetation polygons in a

- Mediterranean test site, ISPRS J. Photogramm. Remote Sens. 63 (2) (2008) 237–250, <https://doi.org/10.1016/j.isprsjprs.2007.08.007>.
- [134] D.A. Aldred, J. Wang, A method for obtaining and applying classification parameters in object-based urban rooftop extraction from VHR multispectral images, Int. J. Remote Sens. 32 (10) (2011) 2811–2823, <https://doi.org/10.1080/01431161003745590>.
- [135] J. Yang, T. Jones, J. Caspersen, Y. He, Object-based canopy gap segmentation and classification: Quantifying the pros and cons of integrating optical and LiDAR data, Remote Sens. 7 (12) (2015) 15917–15932, <https://doi.org/10.3390/rs71215811>.
- [136] J.C. Tilton, Y. Tarabalka, P.M. Montesano, E. Gofman, Best merge region-growing segmentation with integrated nonadjacent region object aggregation, IEEE Trans. Geosci. Remote Sens. 50 (11 PART1) (2012) 4454–4467, <https://doi.org/10.1109/TGRS.2012.2190079>.
- [137] J.C. Tilton, D.K. Hall, G.A. Riggs, Creation of ersatz ground reference data for validating the modis snow and ice product suite, Int. Geosci. Remote Sens. Symp. (2010) 2371–2374, <https://doi.org/10.1109/IGARSS.2010.5650287>.
- [138] J. Damilakis, J. Stratidakis, K. Perisinakis, N. Gourtsoyiannis, Broadband ultrasound attenuation imaging: Algorithm development and clinical assessment of a region growing technique, Phys. Med. Biol. 47 (2) (2002) 315–325, <https://doi.org/10.1088/0031-9155/47/2/310>.
- [139] L. Liu, S. Sclaroff, Region segmentation via deformable model-guided split and merge, Proc. IEEE Int. Conf. Comput. Vis. 1 (2001) 98–104, <https://doi.org/10.1109/iccv.2001.937504>.
- [140] G. GUTMAN, A.C. JANETOS, E.F. MORAN, R.R. RINDFUSS, D. SKOLE, M. A. COCHRANE, *Remote Sensing and Digital Image Processing* (2004).
- [141] B. Guindon, Computer-based aerial image understanding: A review and assessment of its application to planimetric information extraction from very high resolution satellite images, Can. J. Remote Sens. 23 (1) (1997) 38–47, <https://doi.org/10.1080/07038992.1997.10874676>.
- [142] U.C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, M. Heynen, Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information, ISPRS J. Photogramm. Remote Sens. 58 (3–4) (2004) 239–258, <https://doi.org/10.1016/j.isprsjprs.2003.10.002>.
- [143] R. Ohlander, K. Price, D.R. Reddy, Picture segmentation using a recursive region splitting method, Comput. Graph. Image Process 8 (3) (1978) 313–333, [https://doi.org/10.1016/0146-664X\(78\)90060-6](https://doi.org/10.1016/0146-664X(78)90060-6).
- [144] D. Kelkar, S. Gupta, Improved quadtree method for split merge image segmentation, in: *Proc. - 1st Int. Conf. Emerg. Trends Eng. Technol. ICETET 2008*, 2008, pp. 44–47, <https://doi.org/10.1109/ICETET.2008.145>.
- [145] I.N. Manousakas, P.E. Undrill, G.G. Cameron, T.W. Redpath, Split-and-merge segmentation of magnetic resonance medical images: Performance evaluation and extension to three dimensions, Comput. Biomed. Res. 31 (6) (1998) 393–412, <https://doi.org/10.1006/cbmr.1998.1489>.
- [146] H.D. Cheng, X.H. Jiang, Y. Sun, J. Wang, Color image segmentation: Advances and prospects, Pattern Recognit. 34 (12) (2001) 2259–2281, [https://doi.org/10.1016/S0031-3203\(00\)00149-7](https://doi.org/10.1016/S0031-3203(00)00149-7).
- [147] R. Alshehhi, P.R. Marpu, Hierarchical graph-based segmentation for extracting road networks from high-resolution satellite images, ISPRS J. Photogramm. Remote Sens. 126 (May 2020) (2017) 245–260, <https://doi.org/10.1016/j.isprsjprs.2017.02.008>.
- [148] E.H. Al-Hujazi, Wayne, Integration of edge and region basedtechniques for range image segmentation, Intell. Robot. Comput. Vis. 1381 (1991) 589–599.
- [149] J. Le Moigne, J.C. Tilton, S. Member, Refining image segmentation, IEEE Trans. Geosci. Remote Sens. 33 (3) (1995).
- [150] M. Wang, R. Li, Segmentation of high spatial resolution remote sensing imagery based on hard-boundary constraint and two-stage merging, IEEE Trans. Geosci. Remote Sens. 52 (9) (2014) 5712–5725, <https://doi.org/10.1109/TGRS.2013.2292053>.
- [151] R. Gaetano, G. Masi, G. Poggi, L. Verdoliva, G. Scarpa, Marker-controlled watershed-based segmentation of multiresolution remote sensing images, IEEE Trans. Geosci. Remote Sens. 53 (6) (2015) 2987–3004, <https://doi.org/10.1109/TGRS.2014.2367129>.
- [152] X. Li, S.W. Myint, Y. Zhang, C. Galletti, X. Zhang, B.L. Turner, Object-based land-cover classification for metropolitan Phoenix, Arizona, using aerial photography, Int. J. Appl. Earth Obs. Geoinf. 33 (1) (2014) 321–330, <https://doi.org/10.1016/j.jag.2014.04.018>.
- [153] M. Mueller, K. Segl, H. Kaufmann, Edge- and region-based segmentation technique for the extraction of large, man-made objects in high-resolution satellite imagery, Pattern. Recognit. 37 (8) (2004) 1619–1628, <https://doi.org/10.1016/j.patcog.2004.03.001>.
- [154] N. Li, H. Huo, T. Fang, A novel texture-preceded segmentation algorithm for high-resolution imagery, IEEE Trans. Geosci. Remote Sens. 48 (7) (2010) 2818–2828, <https://doi.org/10.1109/TGRS.2010.2041462>.
- [155] B. Johnson, Z. Xie, Unsupervised image segmentation evaluation and refinement using a multi-scale approach, ISPRS J. Photogramm. Remote Sens. 66 (4) (2011) 473–483, <https://doi.org/10.1016/j.isprsjprs.2011.02.006>.
- [156] J. Chen, M. Deng, X. Mei, T. Chen, Q. Shao, L. Hong, Optimal segmentation of a high-resolution remote-sensing image guided by area and boundary, Int. J. Remote Sens. 35 (19) (2014) 6914–6939, <https://doi.org/10.1080/01431161.2014.960617>.
- [157] Z. Wang, C. Lu, X. Yang, Exponentially sampling scale parameters for the efficient segmentation of remote-sensing images, Int. J. Remote Sens. 39 (6) (2018) 1628–1654, <https://doi.org/10.1080/01431161.2017.1410297>.
- [158] A. A. Farag, R. M. Mohamed, and A. El-baz, “Remote Sensing Image Segmentation,” vol. 43, no. 7, pp. 1617–1634, 2005.
- [159] G. Poggi, G. Scarpa, J.B. Zerubia, Supervised segmentation of remote sensing images based on a tree-structured MRF model, IEEE Trans. Geosci. Remote Sens. 43 (8) (2005) 1901–1911, <https://doi.org/10.1109/TGRS.2005.852163>.
- [160] T.N. Tran, R. Wehrens, D.H. Hoekman, L.M.C. Buydens, Initialization of Markov random field clustering of large remote sensing images, IEEE Trans. Geosci. Remote Sens. 43 (8) (2005) 1912–1919, <https://doi.org/10.1109/TGRS.2005.848427>.
- [161] F. Tupin, M. Roux, Markov random field on region adjacency graph for the fusion of SAR and optical data in radargrammetric applications, IEEE Trans. Geosci. Remote Sens. 43 (8) (2005) 1920–1928, <https://doi.org/10.1109/TGRS.2005.852080>.
- [162] C. D'Elia, G. Poggi, G. Scarpa, A tree-structured markov random field model for bayesian image segmentation, IEEE Trans. Image Process. 12 (10) (2003) 1259–1273, <https://doi.org/10.1109/TIP.2003.817257>.
- [163] M.N. Kurnaz, Z. Dokur, T. Ölmez, Segmentation of remote-sensing images by incremental neural network, Pattern Recognit. Lett. 26 (8) (2005) 1096–1104, <https://doi.org/10.1016/j.patrec.2004.10.004>.
- [164] T. Villmann, E. Merényi, B. Hammer, Neural maps in remote sensing image analysis, Neural Networks 16 (3–4) (2003) 389–403, [https://doi.org/10.1016/S0893-6080\(03\)00021-2](https://doi.org/10.1016/S0893-6080(03)00021-2).
- [165] P. Mitra, B. Uma Shankar, S.K. Pal, Segmentation of multispectral remote sensing images using active support vector machines, Pattern Recognit. Lett. 25 (9) (2004) 1067–1074, <https://doi.org/10.1016/j.patrec.2004.03.004>.
- [166] S. Du, Z. Guo, W. Wang, L. Guo, J. Nie, A comparative study of the segmentation of weighted aggregation and multiresolution segmentation, GIScience Remote Sens 53 (5) (2016) 651–670, <https://doi.org/10.1080/15481603.2016.1215769>.
- [167] N. Audebert, B. Le Saux, S. Lefevre, Semantic segmentation of earth observation data using multimodal and multi-scale deep networks, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 10111 LNCS (2017) 180–196, https://doi.org/10.1007/978-3-319-54181-5_12.
- [168] S. Geman, C. Graffigne, Markov Random Field Image Models and their Applications to Computer Vision, in: Proceedings of the International Congress of ..., 1986, pp. 1496–1517 [Online]. Available: <http://www.mathunion.org/ICM/ICM1986.2/Main/icm1986.2.1496.1517.oc.pdf>.
- [169] Z. T., S.-C. Zhu, Image segmentation by data-driven markov chain monte carlo, IEEE Trans. Pattern Anal. Mach. I 24 (5) (2002) 657–673.
- [170] Q. Yu, D.A. Clausi, IRGS: Image segmentation using edge penalties and region growing, IEEE Trans. Pattern Anal. Mach. Intell. 30 (12) (2008) 2126–2139, <https://doi.org/10.1109/TPAMI.2008.15>.
- [171] M. Jung, E.J. Yun, C.S. Kim, Multiresolution approach for texture segmentation using MRF models, Int. Geosci. Remote Sens. Symp. 6 (2005) 3971–3974, <https://doi.org/10.1109/IGARSS.2005.1525782>.
- [172] G. Moser, S.B. Serpico, Contextual high-resolution image classification by Markovian data fusion, adaptive texture extraction, and multiscale segmentation, Int. Geosci. Remote Sens. Symp. (2011) 1155–1158, <https://doi.org/10.1109/IGARSS.2011.6049402>.
- [173] R. Kemker, C. Salvaggio, C. Kanan, Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning, ISPRS J. Photogramm. Remote Sens. 145 (March) (2018) 60–77, <https://doi.org/10.1016/j.isprsjprs.2018.04.014>.
- [174] I. Nigam, C. Huang, D. Ramanan, Ensemble Knowledge Transfer for Semantic Segmentation, in: *Proc. - 2018 IEEE Winter Conf. Appl. Comput. Vision, WACV 2018* 2018-Janua, 2018, pp. 1499–1508, <https://doi.org/10.1109/WACV.2018.00168>.
- [175] L. Zhang, L. Zhang, B. Du, Deep learning for remote sensing data: A technical tutorial on the state of the art, IEEE Geosci. Remote Sens. Mag. 4 (2) (2016) 22–40, <https://doi.org/10.1109/MGRS.2016.2540798>.
- [176] G. Moser, S.B. Serpico, J.A. Benediktsson, Land-cover mapping by markov modeling of spatial-contextual information in very-high-resolution remote sensing images, Proc. IEEE 101 (3) (2013) 631–651, <https://doi.org/10.1109/JPROC.2012.2211551>.
- [177] L. Zhang, K. Jia, X. Li, Q. Yuan, X. Zhao, Multi-scale segmentation approach for object-based land-cover classification using high-resolution imagery, Remote Sens. Lett. 5 (1) (2014) 73–82, <https://doi.org/10.1080/2150740X.2013.875235>.
- [178] C. Burnett, T. Blaschke, A multi-scale segmentation/object relationship modelling methodology for landscape analysis, Ecol. Model. 168 (3) (2003) 233–249, [https://doi.org/10.1016/S0304-3800\(03\)00139-X](https://doi.org/10.1016/S0304-3800(03)00139-X).
- [179] G. Chen, Q. Weng, G.J. Hay, Y. He, Geographic object-based image analysis (GEOBIA): Emerging trends and future opportunities, GIScience Remote Sens 55 (2) (2018) 159–182, <https://doi.org/10.1080/15481603.2018.1426092>.
- [180] Y. Lecun, L. Bottou, Y. Bengio, P. Ha, Gradient-Based Learning Applied to Document Recognition, in: *Proc. IEEE*, 1998, pp. 1–46.
- [181] S. Hochreiter, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, <https://doi.org/10.17582/journal.pjz/2018.50.6.2199.2207>.
- [182] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495, <https://doi.org/10.1109/TPAMI.2016.2644615>.
- [183] I. Goodfellow, et al., Generative adversarial networks, Commun. ACM 63 (11) (2020) 139–144, <https://doi.org/10.1145/3422622>.
- [184] G.E.H. Alex Krizhevsky, Ilya Sutskever, ImageNet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. (2012) 1097–1105, <https://doi.org/10.1201/9781420010749>.
- [185] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: *3rd Int. Conf. Learn. Represent. ICLR2015 - Conf. Track Proc., 2015*, pp. 1–14.

- [186] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem (2016) 770–778, <https://doi.org/10.1109/CVPR.2016.90>.
- [187] S.R. Christian Szegedy1, Wei Liu2, Yangqing Jia1, Pierre Sermanet1, “Going Deeper with Convolutions, in: Proc. IEEE Conf. Comput. Vis. pattern Recognit, 2015, pp. 1–9, <https://doi.org/10.4324/9781410605337-29>.
- [188] A. G. Howard et al., “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” 2017, [Online]. Available: <http://arxiv.org/abs/1704.04861>.
- [189] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua, 2017, pp. 2261–2269, <https://doi.org/10.1109/CVPR.2017.243>.
- [190] E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (4) (2017) 640–651, <https://doi.org/10.1109/TPAMI.2016.2572683>.
- [191] W. Liu, A. Rabinovich, and A. C. Berg, “ParseNet: Looking Wider to See Better,” pp. 1–11, 2015, [Online]. Available: <http://arxiv.org/abs/1506.04579>.
- [192] G. Wang, W. Li, S. Ourselin, T. Vercauteren, Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 10670 LNCS (2018) 178–190, https://doi.org/10.1007/978-3-319-75238-9_16.
- [193] Y. Li, H. Qi, J. Dai, X. Ji, Y. Wei, Fully convolutional instance-aware semantic segmentation, in: Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua, 2017, pp. 4438–4446, <https://doi.org/10.1109/CVPR.2017.472>.
- [194] Y. Yuan, M. Chao, Y.C. Lo, Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance, IEEE Trans. Med. Imaging 36 (9) (2017) 1876–1886, <https://doi.org/10.1109/TMI.2017.2695227>.
- [195] N. Liu, H. Li, M. Zhang, J. Liu, Z. Sun, T. Tan, Accurate iris segmentation in non-cooperative environments using fully convolutional networks, in: 2016 Int. Conf. Biometrics, ICB 2016, 2016, <https://doi.org/10.1109/ICB.2016.7550055>.
- [196] R. Hooda, A. Mittal, S. Sofat, Lung segmentation in chest radiographs using fully convolutional networks, Turkish J. Electr. Eng. Comput. Sci. 27 (2) (2019) 710–722, <https://doi.org/10.3906/elk-1710-157>.
- [197] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2018) 834–848, <https://doi.org/10.1109/TPAMI.2017.2699184>.
- [198] A. G. Schwing and R. Urtasun, “Fully Connected Deep Structured Networks,” pp. 1–10, 2015, [Online]. Available: <http://arxiv.org/abs/1503.02351>.
- [199] S. Zheng, et al., Conditional random fields as recurrent neural networks, Proc. IEEE Int. Conf. Comput. Vis. 2015 Inter (2015) 1529–1537, <https://doi.org/10.1109/ICCV.2015.179>.
- [200] G. Lin, C. Shen, A. Van Den Hengel, I. Reid, Efficient piecewise training of deep structured models for semantic segmentation, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem (2016) 3194–3203, <https://doi.org/10.1109/CVPR.2016.348>.
- [201] Z. Liu, X. Li, P. Luo, C.C. Loy, X. Tang, Semantic image segmentation via deep parsing network, Proc. IEEE Int. Conf. Comput. Vis. 2015 Inter (2015) 1377–1385, <https://doi.org/10.1109/ICCV.2015.162>.
- [202] X. Chen, C.L. Zitnick, Mind’s eye: A recurrent visual representation for image caption generation, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 07-12-June (2015) 2422–2431, <https://doi.org/10.1109/CVPR.2015.7298856>.
- [203] X. Jia, E. Gavves, B. Fernando, T. Tuytelaars, Guiding the long-short term memory model for image caption generation, Proc. IEEE Int. Conf. Comput. Vis. 2015 Inter (2015) 2407–2415, <https://doi.org/10.1109/ICCV.2015.277>.
- [204] J. Donahue, et al., Long-term recurrent convolutional networks for visual recognition and description, IEEE Trans. Pattern Anal. Mach. Intell. 39 (4) (2017) 677–691, <https://doi.org/10.1109/TPAMI.2016.2599174>.
- [205] A. Karpathy, L. Fei-Fei, Deep visual-semantic alignments for generating image descriptions, IEEE Trans. Pattern Anal. Mach. Intell. 39 (4) (2017) 664–676, <https://doi.org/10.1109/TPAMI.2016.2598339>.
- [206] J. Mao, W. Xu, Y. Yang, J. Wang, Z. Huang, A. Yuille, Deep captioning with multimodal recurrent neural networks (m-RNN), in: 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. 1090, 2015, pp. 1–17.
- [207] O. Vinyals, A. Toshev, S. Bengio, D. Erhan, Show and tell: A neural image caption generator, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 07-12-June, 2015, pp. 3156–3164, <https://doi.org/10.1109/CVPR.2015.7298935>.
- [208] K. Xu, et al., Show, attend and tell: Neural image caption generation with visual attention, in: 32nd Int. Conf. Mach. Learn. ICML 2015 3, 2015, pp. 2048–2057.
- [209] Q. You, H. Jin, Z. Wang, C. Fang, J. Luo, Image captioning with semantic attention, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem, 2016, pp. 4651–4659, <https://doi.org/10.1109/CVPR.2016.503>.
- [210] H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, in: Proc. IEEE Int. Conf. Comput. Vis. 2015 Inter, 2015, pp. 1520–1528, <https://doi.org/10.1109/ICCV.2015.178>.
- [211] A. Kendall, V. Badrinarayanan, R. Cipolla, Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding, in: Br. Mach. Vis. Conf. 2017, BMVC 2017, 2017, <https://doi.org/10.5244/c.31.57>.
- [212] Y. Yuan, X. Chen, J. Wang, Object-contextual representations for semantic segmentation, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 12351 LNCS (2020) 173–190, https://doi.org/10.1007/978-3-030-58539-6_11.
- [213] and H.L. Jun Fu, Jing Liu, Yuhang Wang, Stacked Deconvolutional Network for Semantic Segmentation, in: Proc. - Int. Conf. Image Process. ICIP, 2018, pp. 1573–1577, <https://doi.org/10.1109/ICIP.2018.8451256>.
- [214] A. Chaurasia, E. Culurciello, LinkNet: Exploiting encoder representations for efficient semantic segmentation, in: 2017 IEEE Vis. Commun. Image Process. VCIP 2017 2018-Janua, 2018, pp. 1–4, <https://doi.org/10.1109/VCIP.2017.8305148>.
- [215] X. Xia and B. Kulic, “W-Net: A Deep Model for Fully Unsupervised Image Segmentation,” 2017, [Online]. Available: <http://arxiv.org/abs/1711.08506>.
- [216] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” pp. 1–8.
- [217] F. Milletari, N. Navab, S.A. Ahmadi, V-Net: Fully convolutional neural networks for volumetric medical image segmentation, in: Proc. - 2016 4th Int. Conf. 3D Vision, 3DV 2016, 2016, pp. 565–571, <https://doi.org/10.1109/3DV.2016.79>.
- [218] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: Learning dense volumetric segmentation from sparse annotation, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 9901 LNCS (2016) 424–432, https://doi.org/10.1007/978-3-319-46723-8_49.
- [219] Z. Zhou, M.M. Rahman Siddique, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 11045 LNCS (2018) 3–11, https://doi.org/10.1007/978-3-030-00889-5_1.
- [220] Z. Zhang, Q. Liu, Y. Wang, Road extraction by deep residual U-Net, IEEE Geosci. Remote Sens. Lett. 15 (5) (2018) 749–753, <https://doi.org/10.1109/LGRS.2018.2802944>.
- [221] Y. Ding, F. Chen, Y. Zhao, Z. Wu, C. Zhang, D. Wu, A stacked multi-connection simple reducing Net for brain tumor segmentation, IEEE Access. 7 (2019) 104011–104024, <https://doi.org/10.1109/ACCESS.2019.2926448>.
- [222] M. Ghafoorian, et al., Deep multi-scale location-aware 3D convolutional neural networks for automated detection of lacunes of presumed vascular origin, Neuroimage Clin. 14 (2017) 391–399, <https://doi.org/10.1016/j.nic.2017.01.033>.
- [223] A. Youssef, D.D. Bloisi, M. Muscio, A. Pennisi, D. Nardi, A. Facchiano, Deep Convolutional Pixel-wise Labeling for Skin Lesion Image Segmentation, in: MeMeA 2018 - 2018 IEEE Int. Symp. Med. Meas. Appl. Proc., 2018, <https://doi.org/10.1109/MeMeA.2018.8438669>.
- [224] X. Li, T. Lai, S. Wang, Q. Chen, C. Yang, R. Chen, Weighted feature pyramid networks for object detection, in: Proc. - 2019 IEEE Intl Conf Parallel Distrib. Process. with Appl. Big Data Cloud Comput. Sustain. Comput. Commun. Soc. Comput. Networking, ISPA/BDCloud/SustainCom/SocialCom 2019, 2019, pp. 1500–1504, <https://doi.org/10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00217>.
- [225] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua, 2017, pp. 6230–6239, <https://doi.org/10.1109/CVPR.2017.660>.
- [226] G. Ghiasi, C.C. Fowlkes, Laplacian pyramid reconstruction and refinement for semantic segmentation, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 9907 LNCS (2016) 519–534, https://doi.org/10.1007/978-3-319-46487-9_32.
- [227] J. He, Z. Deng, Y. Qiao, Dynamic multi-scale filters for semantic segmentation, Proc. IEEE Int. Conf. Comput. Vis. 2019-Octob (2019) 3561–3571, <https://doi.org/10.1109/ICCV.2019.00366>.
- [228] H. Ding, X. Jiang, B. Shuai, A.Q. Liu, G. Wang, Context Contrasted Feature and Gated Multi-scale Aggregation for Scene Segmentation, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2018, pp. 2393–2402, <https://doi.org/10.1109/CVPR.2018.00254>.
- [229] Z. Liu, X. Dong, J. Xue, H. Li, Y. Chen, Adaptive neural control for a class of pure-feedback nonlinear systems via dynamic surface technique, IEEE Trans. Neural Networks Learn. Syst. 27 (9) (2016) 1969–1975, <https://doi.org/10.1109/TNNLS.2015.2462127>.
- [230] D. Lin, Y. Ji, D. Lischinski, D. Cohen-Or, H. Huang, Multi-scale context intertwining for semantic segmentation, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 11207 LNCS (2018) 622–638, https://doi.org/10.1007/978-3-030-01219-9_37.
- [231] G. Li, P. Yan, Y. Xie, G. Wang, L. Lin, Y. Yu, Instance-level salient object segmentation, Comput. Vis. Image Underst. 207 (March) (2021) 103207, <https://doi.org/10.1016/j.cviu.2021.103207>.
- [232] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (6) (2017) 1137–1149, <https://doi.org/10.1109/TPAMI.2016.2577031>.
- [233] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2) (2020) 386–397, <https://doi.org/10.1109/TPAMI.2018.2844175>.
- [234] S. Liu, L. Qi, H. Qin, J. Shi, J. Jia, Path aggregation network for instance segmentation, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (1) (2018) 8759–8768, <https://doi.org/10.1109/CVPR.2018.00913>.
- [235] J. Dai, K. He, J. Sun, Instance-aware semantic segmentation via multi-task network cascades, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem (2016) 3150–3158, <https://doi.org/10.1109/CVPR.2016.343>.
- [236] R. Hu, P. Dollar, K. He, T. Darrell, R. Girshick, Learning to Segment Every Thing, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2018, pp. 4233–4241, <https://doi.org/10.1109/CVPR.2018.00445>.
- [237] L.C. Chen, A. Hermans, G. Papandreou, F. Schroff, P. Wang, H. Adam, MaskLab: Instance Segmentation by Refining Object Detection with Semantic and Direction

- Features, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2018, pp. 4013–4022, <https://doi.org/10.1109/CVPR.2018.00422>.
- [238] J. Dai, Y. Li, K. He, J. Sun, R-FCN: Object detection via region-based fully convolutional networks, *Adv. Neural Inf. Process. Syst.* (2016) 379–387.
- [239] P.O. Pinheiro, R. Collobert, P. Dollar, Learning to segment object candidates, *Adv. Neural Inf. Process. Syst.* 2015-Janua (2015) 1990–1998.
- [240] E. Xie, et al., PolarMask: Single shot instance segmentation with polar representation, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 1 (2020) 12190–12199, <https://doi.org/10.1109/CVPR42600.2020.01221>.
- [241] X. Guo, X. Lan, K. Wang, S. Li, Contour loss for instance segmentation via k-step distance transformation image, *IET Comput. Vis.* 16 (8) (2022) 683–693, <https://doi.org/10.1049/cvi.12114>.
- [242] Y. Lee, J. Park, Centermask: Real-time anchor-free instance segmentation, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2020, pp. 13903–13912, <https://doi.org/10.1109/CVPR42600.2020.01392>.
- [243] M. Bai, R. Urtasun, Deep watershed transform for instance segmentation, in: *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017* 2017-Janua, 2017, pp. 2858–2866, <https://doi.org/10.1109/CVPR.2017.305>.
- [244] L. Xie, A. Natsev, X. He, J.R. Kender, M. Hill, J.R. Smith, Tracking large-scale video remin in real-world events, *IEEE Trans. Multimed.* 15 (6) (2013) 1244–1254, <https://doi.org/10.1109/TMM.2013.2264929>.
- [245] A. Pathi et al., “Semantic Instance Segmentation via Deep Metric Learning,” 2017, [Online]. Available: <http://arxiv.org/abs/1703.10277>.
- [246] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, in: *4th Int. Conf. Learn. Represent. ICLR2016 - Conf. Track Proc.*, 2016.
- [247] P. Wang, et al., Understanding Convolution for Semantic Segmentation, in: *Proc. - 2018 IEEE Winter Conf. Appl. Comput. WACV 2018* 2018-Janua, 2018, pp. 1451–1460, <https://doi.org/10.1109/WACV.2018.00163>.
- [248] M. Yang, K. Yu, C. Zhang, Z. Li, K. Yang, DenseASPP for semantic segmentation in street scenes, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* (1) (2018) 3684–3692, <https://doi.org/10.1109/CVPR.2018.00388>.
- [249] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, “ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation,” pp. 1–10, 2016, [Online]. Available: <http://arxiv.org/abs/1606.02147>.
- [250] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 11211 LNCS (2018) 833–851, https://doi.org/10.1007/978-3-03-01234-2_49.
- [251] F. Visin, et al., ReSeg: A Recurrent Neural Network-Based Model for Semantic Segmentation, in: *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, 2016, pp. 426–433, <https://doi.org/10.1109/CVPRW.2016.60>.
- [252] F. Visin, K. Kastner, K. Cho, M. Matteucci, A. Courville, and Y. Bengio, “ReNet: A Recurrent Neural Network-Based Alternative to Convolutional Networks,” pp. 1–9, 2015, [Online]. Available: <http://arxiv.org/abs/1505.00393>.
- [253] W. Byeon, T.M. Breuel, F. Raue, M. Litwinski, Scene labeling with LSTM recurrent neural networks, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 07-12-June (2015) 3547–3555, <https://doi.org/10.1109/CVPR.2015.7298977>.
- [254] X. Liang, X. Shen, J. Feng, L. Lin, S. Yan, Semantic object parsing with graph LSTM, in: *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 9905 LNCS, 2016, pp. 125–143, https://doi.org/10.1007/978-3-319-46448-0_8.
- [255] Y. Xiang, D. Fox, DA-RNN: Semantic mapping with data associated recurrent neural networks, *Robot. Sci. Syst.* 13 (2017), <https://doi.org/10.15607/rss.2017.xii.013>.
- [256] R. Hu, M. Rohrbach, T. Darrell, Segmentation from natural language expressions, in: *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 9905 LNCS, 2016, pp. 108–124, https://doi.org/10.1007/978-3-319-46448-0_7.
- [257] M. R. Raza, “Deep Learning-Based Sentiment Analysis For Cloud Provider Selection,” no. March, 2020, <10.13140/RG.2.2.11479.55202>.
- [258] L.C. Chen, Y. Yang, J. Wang, W. Xu, A.L. Yuille, Attention to Scale: Scale-Aware Semantic Image Segmentation, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2016-Decem, 2016, pp. 3640–3649, <https://doi.org/10.1109/CVPR.2016.396>.
- [259] Q. Huang, et al., Semantic segmentation with reverse attention, in: *Br. Mach. Vis. Conf. 2017, BMVC 2017*, 2017, <https://doi.org/10.5244/c.31.18>.
- [260] H. Li, P. Xiong, J. An, L. Wang, Pyramid attention network for semantic segmentation, in: *Br. Mach. Vis. Conf. 2018, BMVC 2018, 2019*, pp. 1–13.
- [261] J. Fu, et al., Dual attention network for scene segmentation, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2019-June, 2019, pp. 3141–3149, <https://doi.org/10.1109/CVPR.2019.00326>.
- [262] Y. Yuan, L. Huang, J. Guo, C. Zhang, X. Chen, and J. Wang, “OCNet: Object Context Network for Scene Parsing,” 2018, [Online]. Available: <http://arxiv.org/abs/1809.00916>.
- [263] X. Li, Z. Zhong, J. Wu, Y. Yang, Z. Lin, H. Liu, Expectation-maximization attention networks for semantic segmentation, in: *Proc. IEEE Int. Conf. Comput. Vis.* 2019-Octob, 2019, pp. 9166–9175, <https://doi.org/10.1109/ICCV.2019.00926>.
- [264] Z. Huang, et al., CCNet: Criss-cross attention for semantic segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 45 (6) (2023) 6896–6908, <https://doi.org/10.1109/TPAMI.2020.3007032>.
- [265] M. Ren, R.S. Zemel, End-to-end instance segmentation with recurrent attention, in: *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017* 2017-Janua, 2017, pp. 293–301, <https://doi.org/10.1109/CVPR.2017.39>.
- [266] H. Zhao, et al., PSANet: Point-wise spatial attention network for scene parsing, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 11213 LNCS (2018) 270–286, https://doi.org/10.1007/978-3-03-01240-3_17.
- [267] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, N. Sang, Learning a discriminative feature network for semantic segmentation, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 1 (2018) 1857–1866, <https://doi.org/10.1109/CVPR.2018.00199>.
- [268] M. Majurski, et al., Cell image segmentation using generative adversarial networks, transfer learning, and augmentations, *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.* 2019-June (2019) 1114–1122, <https://doi.org/10.1109/CVPRW.2019.00145>.
- [269] K. Ehsani, R. Mottaghi, A. Farhad, SeGAN: Segmenting and generating the invisible, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* (2018) 6144–6153, <https://doi.org/10.1109/CVPR.2018.00643>.
- [270] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, “Semantic Segmentation using Adversarial Networks,” 2016, [Online]. Available: <http://arxiv.org/abs/1611.08408>.
- [271] D. Xu, Z. Wang, Semi-supervised semantic segmentation using an improved generative adversarial network, *J. Intell. Fuzzy Syst.* 40 (5) (2021) 9709–9719, <https://doi.org/10.3233/JIFS-202220>.
- [272] W.C. Hung, Y.H. Tsai, Y.T. Liou, Y.Y. Lin, M.H. Yang, Adversarial learning for semi-supervised semantic segmentation, in: *Br. Mach. Vis. Conf. 2018, BMVC 2018*, 2019, pp. 1–17.
- [273] Y. Xue, T. Xu, H. Zhang, R. Long, X. Huang, SegAN:Adversarial network with multi-scale l1 loss for medical image segmentation, *Springer Neuroinformatics c* (2018) 383–392 [Online]. Available: <https://link.springer.com/article/10.1007/s12021-018-9377-x>.
- [274] L. Cai, Y. Chen, N. Cai, W. Cheng, H. Wang, Utilizing amari-alpha divergence to stabilize the training of generative adversarial networks, *Entropy* 22 (4) (2020) 1–19, <https://doi.org/10.3390/E22040410>.
- [275] H. Chen, X. Zhang, Y. Liu, Q. Zeng, Generative adversarial networks capabilities for super-resolution reconstruction of weather radar echo images, *Atmosphere (Basel)* 10 (9) (2019) 1–18, <https://doi.org/10.3390/atmos10090555>.
- [276] Y. Yuan, C. He, Adaptive active contours without edges, *Math. Comput. Model.* 55 (5–6) (2012) 1705–1721, <https://doi.org/10.1016/j.mcm.2011.11.014>.
- [277] X. Chen, B.M. Williams, S.R. Vallabhaneni, G. Czanner, R. Williams, Y. Zheng, Learning active contour models for medical image segmentation, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2019-June, 2019, pp. 11624–11632, <https://doi.org/10.1109/CVPR.2019.01190>.
- [278] T.H.N. Le, K.G. Quach, K. Luu, C.N. Duong, M. Savvides, Reformulating level sets as deep recurrent neural network approach to semantic segmentation, *IEEE Trans. Image Process.* 27 (5) (2018) 2393–2407, <https://doi.org/10.1109/TIP.2018.2794205>.
- [279] C. Rupprecht, E. Huaroc, M. Baust, and N. Navab, “Deep Active Contours,” pp. 1–16, 2016, [Online]. Available: <http://arxiv.org/abs/1607.05074>.
- [280] A. Hatamizadeh, et al., Deep active lesion segmentation, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 11861 LNCS (2019) 98–105, https://doi.org/10.1007/978-3-03-32692-0_12.
- [281] D. Marcos, et al., Learning Deep Structured Active Contours End-to-End, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8877–8885, <https://doi.org/10.1109/CVPR.2018.00925>.
- [282] D. Cheng, R. Liao, S. Fidler, R. Urtasun, Darnet: Deep active ray network for building segmentation, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2019-June, 2019, pp. 7423–7431, <https://doi.org/10.1109/CVPR.2019.00761>.
- [283] A. Hatamizadeh, D. Sengupta, and D. Terzopoulos, “End-to-End Deep Convolutional Active Contours for Image Segmentation,” 2019, [Online]. Available: <http://arxiv.org/abs/1909.13359>.
- [284] H. Zhang, et al., Context Encoding for Semantic Segmentation, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7151–7160, <https://doi.org/10.1109/CVPR.2018.00747>.
- [285] G. Lin, A. Milan, C. Shen, I. Reid, RefineNet: Multi-path refinement networks for high-resolution semantic segmentation, in: *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017* 2017-Janua, 2017, pp. 5168–5177, <https://doi.org/10.1109/CVPR.2017.549>.
- [286] K.M. Lee, H. Myeong, G. Song, SeedNet: Automatic Seed Generation with Deep Reinforcement Learning for Robust Interactive Segmentation, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1760–1768, <https://doi.org/10.1109/CVPR.2018.00189>.
- [287] J. Dai, K. He, J. Sun, BoxSup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation, in: *Proc. IEEE Int. Conf. Comput. Vis.* 2015 Inter, 2015, pp. 1635–1643, <https://doi.org/10.1109/ICCV.2015.191>.
- [288] C. Peng, X. Zhang, G. Yu, G. Luo, J. Sun, Large kernel matters - Improve semantic segmentation by global convolutional network, in: *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017* 2017-Janua, 2017, pp. 1743–1751, <https://doi.org/10.1109/CVPR.2017.189>.
- [289] Z. Wu, C. Shen, A. van den Hengel, Wider or Deeper: Revisiting the ResNet Model for Visual Recognition, *Pattern. Recognit.* 90 (2019) 119–133, <https://doi.org/10.1016/j.patcog.2019.01.006>.
- [290] Z. Zhang, X. Zhang, C. Peng, X. Xue, J. Sun, ExFuse: Enhancing feature fusion for semantic segmentation, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 11214 LNCS (2018) 273–288, https://doi.org/10.1007/978-3-03-01249-6_17.
- [291] M. Mostajabi, P. Yadollahpour, G. Shakhnarovich, Feedforward semantic segmentation with zoom-out features, in: *Proc. IEEE Comput. Soc. Conf. Comput.*

- Vis. Pattern Recognit. 07-12-June, 2015, pp. 3376–3385, <https://doi.org/10.1109/CVPR.2015.7298959>.
- [292] W. Wang, J. Shen, F. Porikli, Saliency-aware geodesic video object segmentation, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 07-12-June, 2015, pp. 3395–3402, <https://doi.org/10.1109/CVPR.2015.7298961>.
- [293] 3 Sun Yat-Sen University 2 The Chinese University of Hong Kong 22*1, 2* Xiaogang Wang, Ping Luo, Guangrui Wang, Liang Lin, Deep dual learning for semantic image segmentation, Nihon Naika Gakkai Zasshi 93 (1) (2004) 161–169, <https://doi.org/10.2169/naika.93.161>.
- [294] X. Li, et al., FoveaNet: Perspective-Aware Urban Scene Parsing, in: Proc. IEEE Int. Conf. Comput. Vis. 2017-Octob, 2017, pp. 784–792, <https://doi.org/10.1109/ICCV.2017.91>.
- [295] I. Kreso, J. Krapac, S. Segvic, Efficient ladder-style densenets for semantic segmentation of large images, IEEE Trans. Intell. Transp. Syst. 22 (8) (2021) 4951–4961, <https://doi.org/10.1109/TITS.2020.2984894>.
- [296] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, N. Sang, BiSeNet: Bilateral segmentation network for real-time semantic segmentation, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 11217 LNCS (2018) 334–349, https://doi.org/10.1007/978-3-030-01261-8_20.
- [297] B. Cheng, et al., SPGNet: Semantic prediction guidance for scene parsing, Proc. IEEE Int. Conf. Comput. Vis. 2019-Octob (2019) 5217–5227, <https://doi.org/10.1109/ICCV.2019.00532>.
- [298] T. Takikawa, D. Acuna, V. Jampani, S. Fidler, Gated-SCNN: Gated shape CNNs for semantic segmentation, Proc. IEEE Int. Conf. Comput. Vis. 2019-Octob (2019) 5228–5237, <https://doi.org/10.1109/ICCV.2019.00533>.
- [299] J. Fu, et al., Adaptive context network for scene parsing, Proc. IEEE Int. Conf. Comput. Vis. 2019-Octob (2019) 6747–6756, <https://doi.org/10.1109/ICCV.2019.00685>.
- [300] X. Liang, E. Xing, H. Zhou, Dynamic-Structured Semantic Propagation Network, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2018, pp. 752–761, <https://doi.org/10.1109/CVPR.2018.00085>.
- [301] X. Liang, Z. Hu, H. Zhang, L. Lin, E.P. Xing, Symbolic graph reasoning meets convolutions, Adv. Neural Inf. Process. Syst. 2018-Decem (NeurIPS) (2018) 1853–1863.
- [302] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, A. Torralba, Scene parsing through ADE20K dataset, in: Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua, 2017, pp. 5122–5130, <https://doi.org/10.1109/CVPR.2017.544>.
- [303] R. Zhang, S. Tang, Y. Zhang, J. Li, S. Yan, Scale-adaptive convolutions for scene parsing, Proc. IEEE Int. Conf. Comput. Vis. 2017-Octob (2017) 2050–2058, <https://doi.org/10.1109/ICCV.2017.224>.
- [304] T. Xiao, Y. Liu, B. Zhou, Y. Jiang, J. Sun, Unified perceptual parsing for scene understanding, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 11209 LNCS (2018) 432–448, https://doi.org/10.1007/978-3-030-01228-1_26.
- [305] B. Zoph, et al., Rethinking pre-training and self-training, Adv. Neural Inf. Process. Syst. 2020-Decem (NeurIPS) (2020).
- [306] X. Zhang, et al., DCNAs: Densely connected neural architecture search for semantic image segmentation, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2021, pp. 13951–13962, <https://doi.org/10.1109/CVPR46437.2021.01374>.
- [307] A. Tao, K. Sapra, and B. Catanzaro, “Hierarchical Multi-Scale Attention for Semantic Segmentation,” pp. 1–11, 2020, [Online]. Available: <http://arxiv.org/abs/2005.10821>.
- [308] A. Kirillov, K. He, R. Girshick, C. Rother, P. Dollar, Panoptic segmentation, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019-June (2019) 9396–9405, <https://doi.org/10.1109/CVPR.2019.00963>.
- [309] A. Kirillov, R. Girshick, K. He, P. Dollar, Panoptic feature pyramid networks, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019-June (2019) 6392–6401, <https://doi.org/10.1109/CVPR.2019.00656>.
- [310] Y. Li, et al., Attention-guided unified network for panoptic segmentation, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019-June (2019) 7019–7028, <https://doi.org/10.1109/CVPR.2019.00719>.
- [311] L. Porzi, S.R. Bulo, A. Colovic, P. Kontschieder, Seamless scene segmentation, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019-June (2019) 8269–8278, <https://doi.org/10.1109/CVPR.2019.00847>.
- [312] B. Cheng et al., “Panoptic-DeepLab,” pp. 1–4, 2019, [Online]. Available: <http://arxiv.org/abs/1910.04751>.
- [313] Y. Xiong, et al., Upsnet: A unified panoptic segmentation network, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019-June (2019) 8810–8818, <https://doi.org/10.1109/CVPR.2019.00902>.
- [314] R. Mohan, A. Valada, EfficientPS: Efficient panoptic segmentation, Int. J. Comput. Vis. 129 (5) (2021) 1551–1579, <https://doi.org/10.1007/s11263-021-01445-z>.
- [315] M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman, The pascal visual object classes (VOC) challenge, Int. J. Comput. Vis. 88 (2) (2010) 303–338, <https://doi.org/10.1007/s11263-009-0275-4>.
- [316] R. Mottaghi, et al., The role of context for object detection and semantic segmentation in the wild, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2014, pp. 891–898, <https://doi.org/10.1109/CVPR.2014.119>.
- [317] T.Y. Lin, et al., Microsoft COCO: Common objects in context, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 8693 LNCS (PART 5) (2014) 740–755, https://doi.org/10.1007/978-3-319-10602-1_48.
- [318] M. Cordts, et al., The Cityscapes Dataset for semantic urban scene understanding, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem (2016) 3213–3223, <https://doi.org/10.1109/CVPR.2016.350>.
- [319] C. Liu, J. Yuen, and A. Torralba, “Nonparametric scene parsing: Label transfer via dense scene alignment,” pp. 1972–1979, 2010, <https://doi.org/10.1109/cvpr.2009.5206536>.
- [320] S. Gould, R. Fulton, D. Koller, Decomposing a scene into geometric and semantically consistent regions, in: Proc. IEEE Int. Conf. Comput. Vis., 2009, pp. 1–8, <https://doi.org/10.1109/ICCV.2009.5459211>.
- [321] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, Proc. IEEE Int. Conf. Comput. Vis. 2 (2001) 416–423, <https://doi.org/10.1109/ICCV.2001.937655>.
- [322] A. Prest et al., “Learning Object Class Detectors from Weakly Annotated Video To cite this version : HAL Id : hal-00695940 Learning Object Class Detectors from Weakly Annotated Video,” 2012.
- [323] A. Geiger, P. Lenz, C. Stiller, R. Urtasun, Vision meets robotics: The KITTI dataset, Int. J. Rob. Res. 32 (11) (2013) 1231–1237, <https://doi.org/10.1177/0278364913491297>.
- [324] J.M. Alvarez, T. Gevers, Y. Lecun, A.M. Lopez, LNCS 7578 - Road scene segmentation from a single image, Eur. Conf. Comput. Vis. (ECCV2012) 7578 (2012) 376–389 [Online]. Available: https://link.springer.com/content/pdf/10.1007%2F978-3-642-33786-4_28.pdf.
- [325] L. Wang et al., “Learning to Detect Salient Objects with Image-level Supervision”.
- [326] D. Batra, D. Parikh, J. Luo, and T. Chen, “iCoseg : Interactive Co-segmentation with Intelligent Scribble Guidance,” 2009.
- [327] N. Silberman and R. Fergus, “Indoor Scene Segmentation using a Structured Light Sensor”.
- [328] V. Gupta and S. Raman, “Automatic Trimap Generation for Image Matting Automatic Trimap Generation for Image,” no. July, 2017.
- [329] P. Luo, X. Wang, X. Tang, Pedestrian parsing via deep decompositional network, in: Proc. IEEE Int. Conf. Comput. Vis., 2013, pp. 2648–2655, <https://doi.org/10.1109/ICCV.2013.329>.
- [330] S. W. Zamir et al., “iSAID: A Large-scale Dataset for Instance Segmentation in Aerial Images,” 2019, [Online]. Available: <http://arxiv.org/abs/1905.12886>.
- [331] W. Cai, K. Jin, and J. Hou, “VDD : Varied Drone Dataset for Semantic Segmentation,” 2023.
- [332] B. Hariharan, P. Arbeláez, L. Bourdev, S. Maji, J. Malik, Semantic contours from inverse detectors, in: Proc. IEEE Int. Conf. Comput. Vis., 2011, pp. 991–998, <https://doi.org/10.1109/ICCV.2011.6126343>.
- [333] X. Chen, R. Mottaghi, X. Liu, S. Fidler, R. Urtasun, A. Yuille, Detect what you can: Detecting and representing objects using holistic models and body parts, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (015) (2014) 1979–1986, <https://doi.org/10.1109/CVPR.2014.254>.
- [334] G. Ros, L. Sellart, J. Materzynska, D. Vazquez, A.M. Lopez, The SYNTHIA Dataset: A large collection of synthetic images for semantic segmentation of urban scenes, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem (600388) (2016) 3234–3243, <https://doi.org/10.1109/CVPR.2016.352>.
- [335] X. Shen, et al., Automatic portrait segmentation for image stylization, Comput. Graph. Forum 35 (2) (2016) 93–102, <https://doi.org/10.1111/cgf.12814>.
- [336] P. Bilic et al., “The Liver Tumor Segmentation Benchmark (LITS)”.
- [337] J. Zhang et al., “Robust Retinal Vessel Segmentation via Locally Adaptive Derivative Frames in Orientation Scores,” vol. 35, no. 12, pp. 2631–2644, 2016.
- [338] N.S.T. Sai, R.C. Patil, Image retrieval using 2D dual-tree discrete wavelet transform, Int. J. Comput. Appl. 14 (6) (2011) 1–8, <https://doi.org/10.5120/1891-2513>.
- [339] Z.H. Zhou, A brief introduction to weakly supervised learning, Natl. Sci. Rev. 5 (1) (2018) 44–53, <https://doi.org/10.1093/nsr/nwx106>.
- [340] L. Jing, Y. Tian, Self-Supervised visual feature learning with deep neural networks: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 43 (11) (2021) 4037–4058, <https://doi.org/10.1109/TPAMI.2020.2992393>.
- [341] R. Anirudh, J.J. Thiagarajan, T. Bremer, H. Kim, Lung nodule detection using 3D convolutional neural networks trained on weakly labeled data, Med. Imaging 2016 Comput. Diagnosis 9785 (2016) 978532, <https://doi.org/10.1117/12.2214876>.
- [342] X. Feng, J. Yang, A.F. Laine, E.D. Angelini, Discriminative localization in CNNs for weakly-supervised segmentation of pulmonary nodules, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 10435 LNCS (2017) 568–576, https://doi.org/10.1007/978-3-319-66179-7_65.
- [343] S. Hwang, H.E. Kim, Self-Transfer learning for weakly supervised lesion localization, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 9901 LNCS (2016) 239–246, https://doi.org/10.1007/978-3-319-46723-8_28.
- [344] J. Kleesiek, et al., Deep MRI brain extraction: A 3D convolutional neural network for skull stripping, Neuroimage 129 (2016) 460–469, <https://doi.org/10.1016/j.neuroimage.2016.01.024>.
- [345] N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi, Medical image computing and computer-Assisted intervention - MICCAI 2015: 18th International Conference Munich, Germany, October 5-9, 2015 proceedings, part III, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 9351 (Cvd) (2015) 12–20, <https://doi.org/10.1007/978-3-319-24574-4>.
- [346] X.Y. Zhou, C. Riga, S.L. Lee, G.Z. Yang, Towards automatic 3D shape instantiation for deployed stent grafts: 2D multiple-class and class-imbalance marker segmentation with equally-weighted focal U-Net, IEEE Int. Conf. Intell. Robot. Syst. (2018) 1261–1267, <https://doi.org/10.1109/IROS.2018.8594178>.
- [347] F. Milletari, et al., Hough-CNN: Deep learning for segmentation of deep brain regions in MRI and ultrasound, Comput. Vis. Image Underst. 164 (2017) 92–102, <https://doi.org/10.1016/j.cviu.2017.04.002>.