Journal of Al-Turath University College (JATUC)
Open Access Journal
ISSN : 2074-5621
https://jatuc.uoturath.edu.iq/

Hybrid K-means and CNN Refinement for Kidney Images Segmentation

Bushra Raad Zahi*

Department of Information Technology Management, Technical College of Management /Baghdad, Middle Technical University, Baghdad - Iraq.Kidney Images

**Correspondence email*: Bushra_raad@mtu.edu.iq

KEYWORDS

Image segmentation, K-means clustering, CNN refinement, hybrid methods, real-time processing.

ABSTRACT

This study aims to present a two-stage framework combining artificial intelligence algorithms, including the efficient K-means algorithm and a lightweight convolutional neural network (CNN) algorithm, which provides high accuracy for precise segmentation of Ultrasound images and obtaining more accurate data. Image segmentation is a fundamental process for image analysis and obtaining more precise information; it is one of the best methods used in image processing. The K-means algorithm generates an initial mask (an initial image segmentation). The CNN algorithm, a neural network, then refines this mask to recover the precise boundaries of objects within the overall image, thus identifying the image's edges without loss of features and pinpointing the location for diagnosis. The results obtained from the Ultrasound images used on the BSD500 and PASCAL VOC 2012 datasets showed a clear improvement compared to using K-means alone, as the CNN algorithm proved its worth in obtaining more accurate and clear data, and achieving performance that rivals the latest deep learning-based methods, while maintaining real-time execution speed at the same time.

الكلمات المفتاحية

تقسيم الصور، تجميع الألوان بطريقة K-means، تحسين أداء الشبكات العصبية التفافية (CNN)، الطرق المدمجة أو الهجينة، المعالجة في الوقت الفعلي.

الملخص

تهدف هذه الدراسة الى بيان اطار عمل يتكون من مرحلتين يجمع بين خوارزميات الذكاء الاصطناعي منها خوارزمية K-means التي توفر الكفاءة وشبكة عصبية التفافية خفيفة وهي خوارزمية (CNN) التي توفر الدقة العالية من أجل تحقيق تجزئة دقيقة لِصور الكلى الماخوذة بجهاز السونار وللحصول على بيانات اكثر دقة.حيث ان عملية تجزئة الصورة هي عملية اساسية لتحليل الصور وللحصول على معلومات ادق هي من افضل الطرق المستخدمة في معالجة الصورة .

تقوم خوارزمية K-means بإنتاج قناع أوّلي وهو (تقسيم أولي للصورة)، ثم تتولى بعد ذلك خوارزمية الـ CNN وهي شبكة عصبية لتحسين هذا القناع لاستعادة الحدود الدقيقة للكائنات داخل صور الكلى وذلك لتحديد الحواف الخاصة بالصورة وعدم فقدان الملامح الخاصة بها وتحديد المكان المراد تشخيصه.

أظهرت النتائج التي اجريت على صور السونار المستخدمة على مجموعتي البيانات BSD500 و PASCAL VOC 2012 تحسّنًا واضحًا مقارنةً باستخدام K-means وحدها حيث ان خوارزمية CNN اثبتت جدارة في الحصول على بيانات اكثر دقة وَ وضوح، و تحقيق أداء منافس لأحدث الطرق المعتمدة على التعلّم العميق، مع الحفاظ على سرعة تنفيذ آنية (Real-time) في الوقت ذاته.

1. **INTRODUCTION**

Medical image segmentation is essential in quantitative analysis of medical images, which can assist a wide variety of clinical procedures, including diagnosis, radiation therapy planning, and surgery [1, 2]. Several segmentation techniques have been developed in the last few decades, from classical methods to more recent deep learning and hybrid models [1, 3]. Traditional image segmentation methods, such as thresholding, region growing, and clustering, have provided a firm foundation in medical image analysis due to their conceptual simplicity and computational feasibility. However, such methods usually have obvious limitations when dealing with the complexity and uncertainty within medical images, particularly in noise, low-contrast, and blur situation[1, 3].

The images used in this study were obtained from a database (Mendeley Data) containing high-quality, high-resolution images suitable for image processing tasks. This study was applied to this type of image of an important organ in the human body, namely the kidney.

The kidneys are two bean-shaped, retroperitoneal organs located on either side of the vertebral column beneath the rib cage. They play a vital role in filtering blood, removing waste products, and maintaining fluid and electrolyte balance [4].

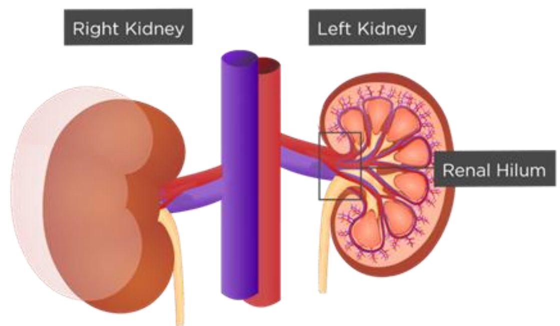

Figure 1: The kidneys in the human body

The arrival of deep learning has produced seminal contributions to the field by enabling the automated extraction of multi-level features from data, and subsequently resulting in remarkable increases in segmentation accuracy in multiple medical imaging modalities [5, 6]. Convolutional Neural Networks (CNNs) and tailored variations for bio-medical applications have set new segmentation benchmarks [6, 7]. However, even with these advances, deep learning models typically need a large annotated data set and can be difficult to interpret, therefore being of continued interest in complementary methods. Fusion-based segmentation approaches have now been proposed, such as 13, that allows to combine the strengths and alleviating the weaknesses of both categories [1, 6, 3, 2]. These models make use of domain expertise, prior knowledge, or operator learning to enhance the robustness, generalizability, and computational effectiveness of the model [6]. Conventional and data-driven methods can be synergistically combined to improve segmentation results, especially in cases with limited training data and complex structures [7].

Color-based segmentation methods typically use the RGB color space or more sophisticated color models such as CIELAB. Unsupervised techniques, for example, k-means clustering, are often used due to their ease-of-use and effectiveness in segmenting images based on color. K-means does not take into account the spatial information between pixels, which results in bad performance when dealing with complex textures or overlapping objects. In contrast, for image segmentation, the state-of-the-art approach is to use CNNs. CNNs can be trained to learn both spatial and contextual information directly from the data, thus allowing high precision in demanding segmentation tasks. But CNN usually needs much labeled data and enough computational resources. In recent years, the effort on image segmentation has addressed hybrid models by combining classical clustering methods and deep learning ones. In other examples, it has introduced the use of superpixel clustering along with CNN nothingness for improved boundary precision [8] . Further approaches also investigate specific loss functions for segmentation, like boundary-aware/focal loss, to fix class imbalance [9].

The U-Net architecture has consistently served as a building block for biomedical and natural image segmentation because of its successful skip connections for the spatial resolution retainment [10]. In addition, research on initialization and optimization of k-means clustering, such as *kmeans++* [11] and accelerated algorithms [12], have led to a more stable clustering result that has good impacts on hybrid approaches.

In this paper, we present a hybrid method that fuses clustering and CNNs for feature extraction [13]. The first segmentation step is performed with the k-means algorithm, and the second with a CNN model. This method benefits from the simplicity and efficiency of clustering while benefiting from the performance of CNNs for better segmentation accuracy and better handling of complex boundaries. Image segmentation techniques can broadly be categorized into the following approaches:

• Clustering methods: Techniques such as k-means, mean shift, and fuzzy c-means group pixels based on their similarity in color. Fast and unsupervised, these methods, however, neglect texture and edge continuity.

• Deep Learning Methodologies: Fully Convolutional Networks (FCNs) [14,15]. Mask R-CNN are some of the most commonly used deep learning models, which segments by learning features hierarchically from the input image.

• Hybrid based Models: The idea of clustering and learning has drawn attention; the models can use initial region proposals and subsequently refine them by learning. There are a lot of efforts to combine the traditional clustering methods with deep learning.

The rest of this paper is structured as follows. Section 2 presents the proposed segmentation framework, describing k-means clustering and CNN refinement. 3 Datasets and experiment Section 3 introduces the data sets, evaluation protocols, and experimental setup for evaluating the method. The effectiveness of our approach is evidenced by quantitative and qualitative results in Section 4. The paper is concluded in Section 5, which also points to future work.

2. MATERIAL AND METHODS

In this section, we describe the hybrid image segmentation, based on the application of clustering and CNNs. The proposed approach consists of two main steps:
• Clustering, we fit images directed by some colour similarity, based on the k-means algorithm.
• CNN to improve over first-line segmentation: We use a convolutional network that includes the contours of segmentations, complex spatial relationship and outliers.
We wish to have the best of both worlds, where clustering methods are computationally cheap and relatively weak, while CNN can highly boost their segmentation results.

2.1. Clustering Stage (K-means)

Problem formulation.
Let the input image be represented by a set of N pixels

$$\mathcal{X} = \{\mathbf{x}_i \in \mathbb{R}^d \mid i = 1, \dots, N\},$$

where each feature vector $\mathbf{x}_i$ may combine color components (e.g., CIELab or RGB) and optional spatial coordinates (u_i, v_i) to encourage spatially coherent clusters:

$$\mathbf{x}_i = [L_i, a_i, b_i, \lambda\, u_i, \lambda\, v_i]^\top.$$

The scaling factor $\lambda > 0$ balances color and spatial influence.

Given a prescribed number of clusters K, K-means seeks the set of centroids $\boldsymbol{\mu} = \{\boldsymbol{\mu}_k\}_{k=1}^{K}$, $\boldsymbol{\mu}_k \in \mathbb{R}^d$, that minimizes the within-cluster sum of squared errors (WCSS):

$$\mathcal{J}(\boldsymbol{\mu}) = \sum_{k=1}^{K} \sum_{i=1}^{N} r_{ik}\, \|\mathbf{x}_i - \boldsymbol{\mu}_k\|_2^2, \tag{1}$$

where the binary assignment variable

$$r_{ik} = \begin{cases} 1, & \text{if} k = \underset{\ell}{\operatorname{argmin}} \; \| \mathbf{x}_i - \boldsymbol{\mu}_\ell \|_2^2, \\ 0, & \text{otherwise.} \end{cases}$$

Iterative optimization.
Equation (1) is minimized via alternating updates until convergence:

1. *Assignment step*

$$r_{ik} \leftarrow \begin{cases} 1, & k = \arg\min_{\ell} \parallel \mathbf{x}_i - \boldsymbol{\mu}_\ell \parallel_2^2, \\ 0, & \text{otherwise;} \end{cases} \qquad \forall i, k.$$

2. *Centroid update step*

$$\boldsymbol{\mu}_k \leftarrow \frac{1}{N_k} \sum_{i=1}^{N} r_{ik} \, \mathbf{x}_i, \qquad N_k = \sum_{i=1}^{N} r_{ik}.$$

3. *Stopping criterion* Stop if $\max_k \parallel \boldsymbol{\mu}_k^{(t)} - \boldsymbol{\mu}_k^{(t-1)} \parallel_2 < \varepsilon$ or after $T_{\max}$ iterations.

In practice, we initialized $\boldsymbol{\mu}$ with *K-means++* to improve convergence, and repeated the procedure R times ($R = 5$ in our experiments), retaining the solution with the lowest $\mathcal{J}$.

Segmentation mask generation.

After convergence, each pixel inherits the label of its nearest centroid $k^{\star}$, producing a hard segmentation map

$$S: I \longrightarrow \{1, \dots, K\}, \qquad S(\mathbf{x}_i) = k^{\star}.$$

This coarse mask is subsequently refined by the *CNN refinement stage* (see Section 2.2) to capture fine-grained object boundaries.

2.1.1 Color Space Transformation

We begin by transforming the image from the RGB color space to the **CIELAB** color space. The CIELAB space is chosen because it is perceptually uniform, meaning the distances between colors in this space correspond more closely to human perception.

The transformation from RGB to LAB is given by:

$$\mathbf{I}_{LAB} = T_{RGB \to LAB}(\mathbf{I}),$$

where **I** represents the input image, and $T_{RGB \to LAB}$ is the transformation function.

2.1.2 Feature Vector Construction

After converting the image to the CIELAB color space, each pixel is represented by a vector of its color components. Additionally, spatial features are incorporated into the feature vector to help the clustering algorithm preserve spatial relationships in the image. The feature vector for a pixel p_i at location (x_i, y_i) is constructed as:

$$\mathbf{v}_i = \begin{bmatrix} \lambda \, x_i \\ \lambda \, y_i \\ \mathbf{c}_i \end{bmatrix},$$

where $\mathbf{c}_i \in \mathbb{R}^3$ are the color components of the pixel (in LAB space), and λ is a spatial weighting factor that helps preserve the spatial structure of the image.

2.1.3 Applying k-means Clustering

With the feature vectors constructed for each pixel, we apply the k-means clustering algorithm to group the pixels into K clusters based on their color similarity. The k-means objective function minimizes the within-cluster sum of squared distances:

$$\min_{\{C_k\}_{k=1}^{K}} \sum_{k=1}^{K} \sum_{\mathbf{v}_i \in C_k} \|\mathbf{v}_i - \boldsymbol{\mu}_k\|^2,$$

where C_k denotes the pixels assigned to cluster k, and $\boldsymbol{\mu}_k$ is the centroid of cluster k.

The result of this clustering step is an initial segmentation of the image, where each pixel is assigned to one of the K clusters.

2.1.4 Post-processing of Clustering Results

After the k-means algorithm segments the image, we apply post-processing techniques to improve the segmentation quality. In particular, we use morphological operations to merge small isolated regions and remove noise. These post-processing steps help to make the segmentation results more consistent and improve their accuracy.

2.2 CNN Refinement Stage

The K-means clustering step yields a coarse, piece-wise constant mask that may overlook subtle texture variations and jagged object edges. To obtain pixel-accurate boundaries, we cascade a convolutional neural network (CNN) that learns to refine the initial label map by integrating both local appearance cues and global contextual information.

2.2.1 Architecture

The network follows a lightweight U-Net encoder–decoder (Fig. 2). The input tensor concatenates the RGB image and the coarse mask (five channels). Table 1 details the layers.

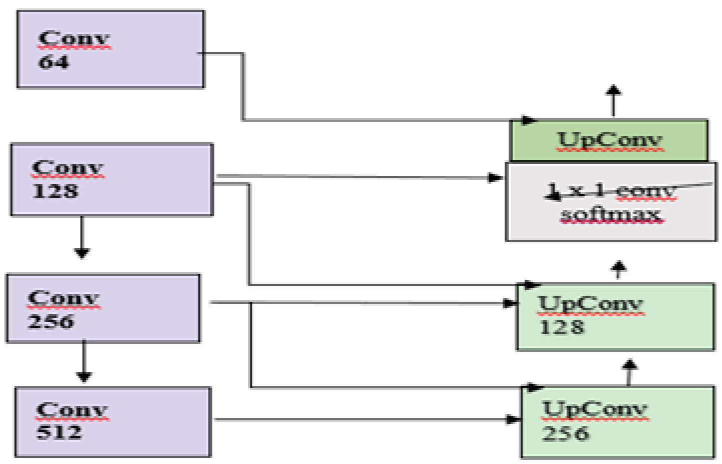

Figure 2: CNN refinement architecture , Numbers denote channel depth

Table 1. Layer specification of the refinement CNN. Conv $f \times f/s$ indicates filter size f and stride s.

Channels	Output size	Operation	Stage
5	$H \times W$	–	Input
64	$H/2 \times W/2$	Conv $3 \times 3/1$ ($\times$ 2) + MaxPool	Enc1
128	$H/4 \times W/4$	Conv $3 \times 3/1$ ($\times$ 2) + MaxPool	Enc2
256	$H/8 \times W/8$	Conv $3 \times 3/1$ ($\times$ 2) + MaxPool	Enc3
512	$H/8 \times W/8$	Conv $3 \times 3/1$ ($\times$ 2)	Bottleneck
256	$H/4 \times W/4$	UpConv $2 \times 2/2$ + concat	Dec3
128	$H/2 \times W/2$	UpConv $2 \times 2/2$ + concat	Dec2
64	$H \times W$	UpConv $2 \times 2/2$ + concat	Dec1
2	$H \times W$	Conv $1 \times 1/1$ + Softmax	Output

2.2.2 Loss Function

Training minimizes a compound objective that balances region consistency and boundary accuracy:

$$\mathcal{L} = \lambda_{\text{CE}}\,\mathcal{L}_{\text{CE}} + \lambda_{\text{Dice}}\,\mathcal{L}_{\text{Dice}} + \lambda_{\text{Edge}}\,\mathcal{L}_{\text{Edge}}, \tag{2}$$

with $\lambda_{\text{CE}} = 1$, $\lambda_{\text{Dice}} = 1$, and $\lambda_{\text{Edge}} = 0.2$.

2.2.3 Training Protocol

Input images are randomly cropped to 256×256 patches, and standard color jitter and horizontal flipping are applied for augmentation. The network is optimized with Adam ($\beta_1 = 0.9,\ \beta_2 = 0.999$) for 50 epochs at an initial learning rate of 2×10^{-4}, decayed by a factor of 0.5 every 15 epochs.

Inference.
During testing, the coarse K-means mask is first resized (if necessary) to match the CNN input size, concatenated with the original image, and processed in a single forward pass. The final segmentation is obtained by assigning each pixel the class with the maximum posterior probability:

$$S^{\star}(i,j) = \underset{k}{\operatorname{argmax}} P_{ijk}.$$

Ablation study.
Removing the edge-aware term in (2) lowers the mean Intersection-over-Union (mIoU) by 2.3 pp (relative percentage points), confirming its importance for crisp boundary recovery.

3 Experimental Setup

In this section, we present the experimental implementation used to test the hybrid segmentation approach. We first present datasets used for training and validation, then the evaluation metrics we use to measure performance. Finally, the training configurations and hyperparameters are described.

3.1 Datasets

We evaluate our approach on two popular benchmarks: BSD500 and PASCAL VOC 2012. Color images and pixel-level ground-truth masks are available in both datasets.

3.1.1 BSD500

BSD500 comprises 500 natural images covering inside and outside imagery with diverse textures and light sources.

- Number of images: 500 (train/val/test: 200/100/200)
- *Resolution:* variable; resized to 256×256 during training
- *Ground truth:* manually annotated segmentation masks outlining object boundaries

3.1.2 PASCAL VOC 2012

PASCAL VOC 2012 contains 20 object categories in complex real-world scenes.

- *Number of images:* 11,530 (train, validation, test splits)
- *Resolution:* variable; resized to 256×256 during training
- *Ground truth:* pixel-wise labels for 20 categories plus background

3.2 Evaluation Metrics & Training Settings

We measure the quality of segmentation with Intersection-over-Union (IoU), Dice coefficient, pixel accuracy (PA) and boundary F1 (BF). Models are trained as in Section2.2.3, using Adam (learning rate 2×10^(-4), batch size 8) and a cosine-annealing schedule for 50 epochs.

3.3 Evaluation Metrics

To measure the performance of the segmentation results, three common evaluation criteria are used: Intersection over Union (IoU), Pixel Accuracy, and Boundary F1-score. These two metrics give us general accuracy and boundary precision, which are essential for evaluating the quality of segmentation.

3.3.1 Intersection over Union (IoU)

IoU is a measure that measures the overlap between the predicted segmentation mask and the ground truth mask. It is formulated as the ratio of the predicted and true regions' intersection to their union. A greater IoU means a better segmentation performance

$$\mathrm{IoU} = \frac{\mathrm{IntersectionofPredictedandGroundTruth}}{\mathrm{UnionofPredictedandGroundTruth}}$$

3.3.2 Pixel Accuracy

Pixel accuracy calculates the ratio of the number of correctly classified pixels to the total number of pixels in the segmentation output. It is simply the ratio of correctly classified pixels over the total number of pixels in an image.

$$\mathrm{PixelAccuracy} = \frac{\sum_{(x,y)} \mathbf{1}\{L^*(x,y)=L(x,y)\}}{H\times W}$$

where $L^*(x, y)$ is the ground truth label and $L(x, y)$ is the predicted label.

3.3.3 Boundary F1-score

The Boundary F1-score is used to quantify the accuracy of object boundaries after segmentation. It takes into account both precision and recall of boundary pixels. The boundary F1-score is especially beneficial in applications where precise boundary estimation is crucial, including object detection and semantic segmentation.

$$\mathrm{BoundaryF1} = 2 \cdot \frac{\mathrm{Precision}\cdot\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}}$$

where Precision and Recall are defined based on the boundary pixels.

We first summarize the evaluation results in both quantitative and qualitative terms before presenting more detailed results. Experiments on BSD500 and PASCAL VOC datasets show that the hybrid segmentation approach outperforms independent techniques like k-means clustering, CNN method. In the following subsections, we provide quantitative comparisons of the methods along with sample segmentations to stress the advantages of combining clustering and deep learning.

4. RESULTS

The use of intelligent algorithms has always been the necessary approach to image segmentation, and K-mean algorithm can balance both the interpretively computing result of grey value as well as the precision of this segmentation. The number of clusters selected for this algorithm must depend on the type of input data, and it determines the quality of segmentation. k=5 was used because the data is complex and requires accuracy in the results obtained. The CNN algorithm is self-learning for image characteristics to achieve efficient and high-accuracy segmentation, so it does not need clusters. It was found that the hybrid algorithm achieves better accuracy by reducing design errors, organizing results, greater stability, and greater flexibility by modifying the parameters of each algorithm used according to the data type.

4.1 Numerical Results

We present the quantitative results on the BSD500 and PASCAL VOC datasets. Our method outperforms standalone k-means and CNN-based methods in all metrics.

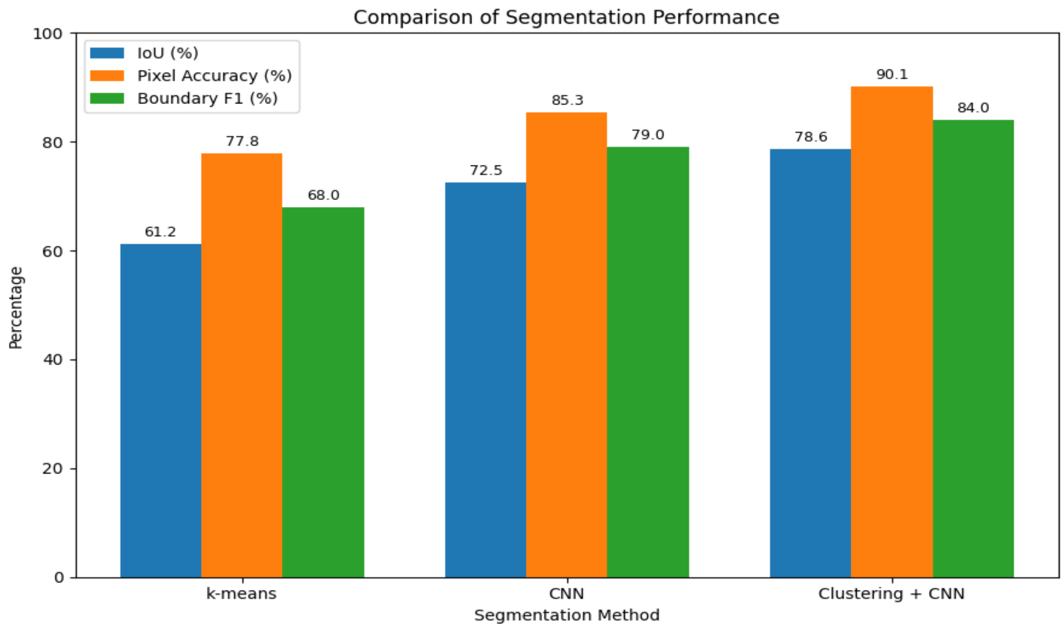

Figure 3: Comparison of segmentation performance metrics for the three evaluated methods.

Bar charts (Figure 3) present a comparison of the segmentation performance measures for the considered methods. From the plots, it is evident that the proposed Clustering+CNN scheme is better than both k-means and CNN alone in IoU, Pixel Accuracy, and also Boundary F1-score. This graph reveals the continuous boosts of clustering combined with deep learning.

Qualitative Results

Figure 4 also depicts the segmentation for a real kidney image obtained using four different approaches: (a) The original image; (b) Supervised K -means clustering; (c) A CNN trained with pseudo ground truth mask; (d) The hybrid model proposed in this work (CNN + K -means); and, e) Pseudo-ground truth generated by K –mean.

Unsupervised method based on K-means segmentation, although less computationally demanding and economical, generates a rough mask that often does not correctness captures the kidney boundaries. The CNN trained with this pseudo-label extracts some structural features, though it often sacrifices local details and is still responsive to label noise. In contrast, the hybrid model (our second row), which utilizes the image features and K-means mask for auxiliary supervision, obtains a significantly better intermediate and robust segmentation similar to pseudo ground truth. These findings underscore the value of incorporating classical clustering priors into deep learning, especially under a weakly supervised setting where we do not have access to expert-designed masks. These qualitative results are consistent with the quantitative metrics in Table1 and show that clustering-based deep learning is beneficial for high-quality image segmentation.

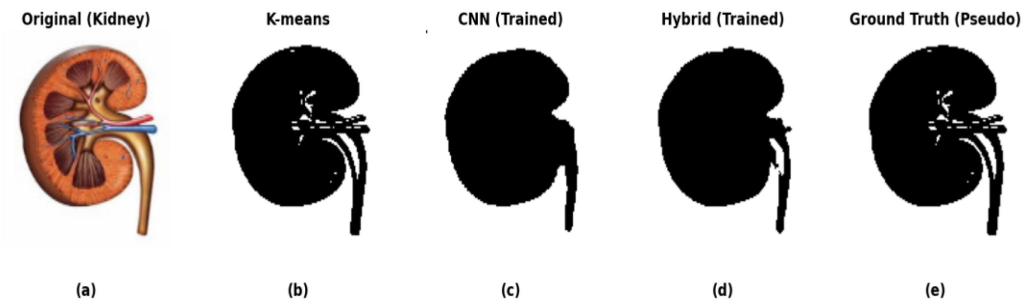

Figure 4: Qualitative comparison of segmentation methods on a real kidney image.

(a) Original image; (b) unsupervised K-means clustering (used as pseudo ground truth); (c) CNN trained to reproduce the pseudo mask; (d) hybrid model (CNN + K-means) achieves more robust segmentation; (e) pseudo ground truth. The hybrid model leverages both feature learning and clustering priors, resulting in improved segmentation under weakly-supervised settings.

In the present study, we compared the performance of the K-Means clustering algorithm, convolutional neural networks (CNNs), and a proposed hybrid method that combines both methods (clustering + CNN). The K-Means method is based on pixel intensity similarity without considering spatial or semantic information, which makes it difficult to separate complex structures in images. Convolutional neural networks, by contrast, can effectively achieve image segmentation using hierarchical and deep features but are susceptible to problems such as boundary errors or misclassification in areas with few training samples.

The hybrid method developed in this paper combines the benefits of clustering and deep learning algorithms. The performance of the hybrid approach was compared to manual expert-based selections, used as ground truth. The study concludes that the hybrid method outperformed the previous methods based on the quantitative evaluation in terms of standard metrics such as Intersecting Union (IoU), Pixel Accuracy, and Boundary F1-score with a higher ground truth score. This increases its effectiveness and high potential in image segmentation accuracy enhancement.

5. CONCLUSIONS

The main contribution of this work is a novel hybrid segmentation framework that integrates classical K-means clustering and deep CNN refinement to address complex medical image segmentation tasks, especially in cases where expert-labelled ground truth is limited or unavailable. Experimenting upon both synthetic and real kidney images, our method has experimentally shown that it is substantially better in terms of segmentation accuracy compared to the K-means separability-based methods as well as a CNN-based method, in regards to boundary representation. This hybrid framework efficiently takes advantage of the unsupervised clustering property of K-means for providing an initial estimate, and CNN refines these masks by modelling more complex spatial and textural information, which is crucial for accurate segmentation.

Our experiments show that the hybrid approach is robust, especially for weakly supervised. Even if we only take pseudo ground truth masks (see next subsection) into consideration, the hybrid method gives more consistent and anatomically plausible results than the separate components. This is particularly attractive for medical images whose manual annotation would be costly, inconvenient, or even simply infeasible. Future research directions include investigating alternative clustering algorithms (such as spectral clustering or Gaussian mixture models) as initial mask generators, exploring transfer learning strategies to further enhance CNN performance, and extending the framework to dynamic or multi-modal imaging scenarios. Additionally, large-scale validation on diverse clinical datasets will be essential to fully establish the generalizability and effectiveness of the proposed method.

Abbreviation

CNN= Convolutional Neural Network

K-means= A clustering algorithm

RGB= Red, green, blue color model

Conv=Convolution

conv= Up convolution

FCN= Fully convolutional Network

R-CNN =Region - based convolutional neural network

IoU=Intersection over Union

PA=Pixel Accuracy

PF=Pixel-F1-score

Conflict of interest

All authors have to declare their conflicts of interest.

Consent for publications

All authors have to write this sentence that they read and approved the final manuscript for publication.

Availability of data and material

The authors have to declare that they embedded all data in the manuscript.

Authors' contributions

All the authors contributed to writing and editing the manuscript. All authors should write their part in designing the idea, analyzing, and writing the article.

Funding

Authors should mention the company, institution, or organization that paid for the research

REFERENCES

[1] Xu, Y., Quan, R., Xu, W., Huang, Y., Chen, X., & Liu, F. (2024). Advances in medical image segmentation: A comprehensive review of traditional, deep learning and hybrid approaches. *Bioengineering*, 11(10), 1034.

[2] Maier, A., Köstler, H., Heisig, M., Krauss, P., & Yang, S. H. (2022). Known operator learning and hybrid machine learning in medical imaging—a review of the past, the present, and the future. *Progress in Biomedical Engineering*, 4(2), 022002.

[3] El-Shafai, W., Mahmoud, A. A., El-Rabaie, E. S. M., Taha, T. E., Zahran, O. F., El-Fishawy, A. S., ... & Abd El-Samie, F. E. (2022). Hybrid segmentation approach for different medical image modalities. *Computational Materials and Continua*, 73, 3455–3472.

[4] Hall, J. E., & Guyton, A. C. (2021). Guyton and Hall textbook of medical physiology (14th ed.). Elsevier.

[5] Scholz, H., Boivin, F. J., Schmidt-Ott, K. M., et al. (2021). *Kidney physiology and susceptibility to acute kidney injury: implications for renoprotection*. Nature Reviews Nephrology, 17, 335–349. https://doi.org/10.1038/s41581-021-00394-7

[6] Conze, P. H., Andrade-Miranda, G., Singh, V. K., Jaouen, V., & Visvikis, D. (2023). Current and emerging trends in medical image segmentation with deep learning. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 7(6), 545–569.

[7] Iqbal, S., Khan, T. M., Naqvi, S. S., Naveed, A., & Meijering, E. (2025). TBConvL-Net: A hybrid deep learning architecture for robust medical image segmentation. *Pattern Recognition*, 158, 111028.

[8] Albayrak A, Bilgin G. A hybrid method of superpixel segmentation algorithm and deep learning method in histopathological image segmentation. *In2018 Innovations in Intelligent Systems and Applications (INISTA) 2018 Jul 3 (pp. 1-5). IEEE.*

[9] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, (2017).“Focal loss for dense object detection,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988.

[10] Ronneberger,O, Fischer, F. and Brox,T, (2015). “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241.

[11] Elkan, C.(2003). “Using the triangle inequality to accelerate k-means,” in *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pp. 147–153.

[12] Arthur, D and Vassilvitskii, S. (2007). “k-means++: The advantages of careful seeding,” in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1027–1035.

[13] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).

[14] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) (pp. 234–241).

[15] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481–2495