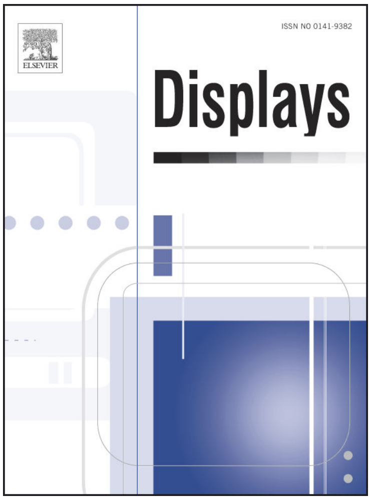


LAIU-Net: A learning-to-augment incorporated robust U-Net for depressed humans' tongue segmentation

Mahmoud Marhamati, Ali Asghar Latifi Zadeh, Masoud Mojdehi Fard, Mohammad Arafat Hussain, Khalegh Jafarnezhad, Ahad Jafarnezhad, Mehdi Bakhtoor, Mohammad Momeny

PII: S0141-9382(23)00004-5
DOI: https://doi.org/10.1016/j.displa.2023.102371
Reference: DISPLA 102371

To appear in: *Displays*

Received Date: 1 July 2022
Revised Date: 18 December 2022
Accepted Date: 5 January 2023

LAIU-Net: A learning-to-augment incorporated robust U-Net for depressed humans' tongue segmentation

Mahmoud Marhamati[1], Ali Asghar Latifi Zadeh[2], Masoud Mojdehi Fard[3], Mohammad Arafat Hussain[4], Khalegh Jafarnezhad[1], Ahad Jafarnezhad[1], Mehdi Bakhtoor[5], Mohammad Momeny[2]*

[1]Esfarayen Faculty of Medical Science, Esfarayen, Iran
[2]Yazd University, Yazd, Iran
[3]Psychosomatic research center, Imam Khomeini Hospital, Tehran University of Medical Sciences, Tehran, Iran
[4]Boston Children's Hospital, Boston, MA 02115
[5]Islamic Azad University, Shirvan, Iran
*Email of Corresponding Author:

mohamad.momeny@gmail.com (M. Momeny)

LAIU-Net: A learning-to-augment incorporated robust U-Net for depressed humans' tongue segmentation

Mahmoud Marhamati[1], Ali Asghar Latifi Zadeh[2], Mohammad Arafat Hussain[3], Khalegh Jafarnezhad[1], Ahad Jafarnezhad[1], Masoud Mojdehi Fard[1], Mohammad Momeny[2]*

[1]Esfarayen Faculty of Medical Science, Esfarayen, Iran
[2]Yazd University, Yazd, Iran
[3]Boston Children's Hospital, Boston, MA 02115
[4]Islamic Azad University, Shirvan, Iran
*Email of Corresponding Author:

mohamad.momeny@gmail.com (M. Momeny)

Abstract

Computer-aided tongue diagnosis system requires segmentation of the tongue body. The frequent movement of the tongue due to its natural flexibility often causes shape variability in photographs across subjects, which makes segmenting the tongue challenging from non-tongue elements, such as the lips, teeth, and other objects in the background of the tongue. The flexibility of the tongue causes a further challenge in maintaining a similar shape and style when taking photos of many healthy subjects and patients. To address these challenges, we have built a tongue dataset, where the tongue of each subject has been scanned thrice with an interval of less than a second. We have collected 333 tongue images from 111 depressed humans, who have been diagnosed with depression by a psychiatrist. In addition, in this paper, we propose a learning-to-augment incorporated U-Net (LAIU-Net) for the segmentation of the depressed human tongue in photographic images. The best policies for data augmentation were automatically chosen with the proposed LAIU-Net. For this purpose, we corrupted photographic tongue images with the Gaussian, speckle, and Poisson noise. The proposed approach addresses the overfitting problem as well as increases the generalizability of a deep network. We have compared the performance of the proposed LAIU-Net with that of other state-of-the-art U-Net configurations. Our LAIU-Net approach achieved a mean boundary F1 score of 93.1%.

Keywords: Tongue segmentation, learning-to-augment strategy, data augmentation, deep learning, U-Net.

1. Introduction

1.1. Tongue Diagnosis

Tongue body segmentation in photographic images is one of the major steps in the computer-aided complete tongue diagnosis system [1]. Previous studies indicated a link between a human tongue and health conditions (e.g., [2]), and diagnosis via examining a tongue has been a common practice worldwide due to its non-painful examination feasibility and obstruction-less accessibility. Human tongue examination has been considered an essential practice to get an insight into the human physiological and pathological conditions in traditional and alternative Chinese medicine, as traditional Asian medicines, are often based on holistic concepts [3]. However, the level of accuracy in tongue examination-based diagnosis varies depending on the experience of clinicians. To improve diagnostic accuracy, clinicians often use other diagnosis approaches (e.g., palpation) in addition to tongue observation. Recently, photographic image-based tongue diagnosis using image processing techniques has been used to address the limitations of the manual tongue-examination approach [4-9]. Chiu proposed a computerized tongue examination system (CTES) based on chromatic and textural analysis strategy [10]. Zhang et al. proposed a Bayesian network-based computer-aided tongue diagnosis system (CATDS) to show the relationship between diseases and tongue-based quantitative attributes [11].

1.2. Tongue Segmentation

The tongue apex (i.e., the front part of the tongue) is very flexible and its movements during photo shooting cause variability in tongue shapes across subjects. This variability causes challenges in segmenting the tongue in the photographic images from non-tongue elements (i.e., lips, teeth, soft and hard palates, etc.) [12]. Segmenting the tongue from the background is essential for tongue-specific feature extraction in a computer-aided tongue diagnosis system. Therefore, the tongue segmentation procedure must be robust across all patients. Typically, photographic images of tongues are captured in color and many traditional image processing-based tongue segmentation approaches used the sequential application of low-level color pixel processing. For example, combined region and edge-based [13], combined region and intensity thresholding-based [13], bi-elliptical deformable contour-based [12], a combination of polar-edge detection and active contour-based [14], color active contour-based [15, 16], color control-geometric and gradient flow snake-based [17], polar edge detection via snake-based [18], combined mean shift algorithm and Canny edge detector-based [19], combined gradient flow and region merging-based [20], threshold control function-based [21, 22], double geodesic flow-based [23], double geo-vector flow-based [24], combined tongue shape and snake correction model-based [25], combined 2D Gabor filter and fast marching-based [26], and adaptive active contour-based [27] tongue segmentation approaches have been proposed in the literature. However, these conventional model-based approaches have various limitations, such as requiring user interaction, sensitivity to parameter settings, etc. [28].

To overcome the limitations of conventional model-based segmentation approaches, supervised learning has been extensively used for medical image segmentation in recent years. Supervised learning-based tongue body segmentation approaches are also reported using support vector machine [29], AdaBoost algorithm [30], cascaded convolutional neural networks (CNN) [31], ResNet [32], SegNet [33], fully convolutional network [34, 35], U-Net [36-38], iterative transfer learning [39], feature pyramid network [40], and patch-driven sparse representation [41]. These supervised learning approaches showed promising tongue segmentation performance by emphasizing producing more accurate segmentation. However, mostly ignored improving the model generalizability aspect of deep models [42-50].

1.3. Proposed model

Our proposed study focuses on improving the generalizability of a supervised deep learning model for tongue segmentation as well as increasing a deep model's robustness to noise. Our approach optimizes parameters of different types of noises via learning to generate diverse augmented data, which in turn improves the generalizability of a deep segmentation model. The contributions of this paper are summarized as follows:

1. To our knowledge, we propose the first learning-to-augment strategy- [51, 52] incorporated robust and generalized deep neural network (i.e., U-Net) for photographic image segmentation. This learning-to-augment strategy follows a computation pipeline, which is well optimized and showed great promise in the classification tasks [51, 52].
2. Our learning-to-augment strategy uses the Bayesian optimizer to choose the best set of policies for data augmentation by selecting the optimized Gaussian, speckle, and Poisson noise parameters.
3. We show the efficacy of our proposed learning-to-augment strategy-incorporated U-Net on the depressed human's tongue segmentation task and demonstrated the best performance compared to other segmentation approaches.
4. We build a photographic tongue dataset consisting of 333 human tongue images, captured from 111 patients diagnosed with depression. Three images were captured in succession from each patient with an interval of less than a second.

2. Materials and Methods

2.1. Depressed Humans' Tongue Dataset

We have collected 333 photographs of the 111 depressed humans' tongues (Aged: 18-60 years, Male: 53, and Female: 58; 3 photos from each patient). Apart from the fact that tongue images are collected from different patients or healthy subjects, many other factors (e.g., the illumination of the photo studio, the color of the food eaten just before the photography, etc.) are responsible for causing substantial variability in the tongue photographs. To take the advantage of these factors to ensure diversity in our dataset, we avoided direct sunlight and took photos of all the patients inside a room using a typical camera flash. We used Canon EOS 80D Digital Camera with an 18-55mm IS STM lens. We set the image resolution to 6000×4000×3 pixels. Table 1 summarizes the camera specification and imaging settings.

Ground truth tongue masks were generated by manual segmentation using the Lasso tools (i.e., standard, polygonal, and magnetic) of Photoshop CC 2019 (Adobe Inc., San Jose, CA)

Table 1. The camera specifications and photography settings

Property	Value	Property	Value
Camera maker	Canon	Dimensions	6000×4000 pixels
Camera model	Canon EOS 80D	Horizontal resolution	72 dpi
F-stop	f/5	Vertical resolution	72 dpi
Exposure time	1/60 sec.	Bit depth	24
ISO speed	ISO-200	Resolution unit	2
Exposure bias	0 step	Color representation	sRGB
Focal length	50mm	Exposure program	Normal
Metering mode	Pattern	EXIF version	0230
Flash mode	Flash, compulsory	White balance	Auto

The tongue apex (i.e., the frontal part of a human tongue) is a very flexible and mobile part of this organ. The tongue apex is followed by the body of the tongue, which is formed by the intrinsic muscles consisting of superior and inferior longitudinal lingual muscles, vertical lingual muscles, and transverse lingual muscles. The body of the tongue is followed by the tongue muscles (also known as extrinsic muscles) with bony insertions. The extrinsic muscles include the genioglossus, the hyoglossus, the styloglossus, and the palatoglossus. The contraction of the geniohyoid against the hyoid bone causes the protrusion of the tongue. The synergy between the intrinsic and extrinsic muscles causes flexibility and mobility of the tongue [53]. We show images of three faces with tongues out of the oral cavity (i.e., out of the mouth) in Figure 1. We can see in this figure that there is substantial variability in the shape of the tongue among these patients. Due to the extensive flexibility of the tongue, maintaining a similar shape of the tongue in photos across patients is difficult. That is why we took three photos of the same tongue in succession with an interval of less than a second. Figure 2 compares our proposed tongue dataset attribute to that of the datasets by Xue et al. [34], Pang et al. [54], Zhou et al. [36], and Huang et al. [37]. We see in this figure that, unlike other datasets, our dataset contains three successive tongue images per patient that in turn allows capturing different shapes of a particular tongue. Photos of the tongue of depressed patients are captured in the psychiatrist's office once a subject is diagnosed with depression. A clinical ethics board of the Islamic Republic of Iran [55] approved this data collection. We also collected informed consent from patients before capturing their tongue photos.

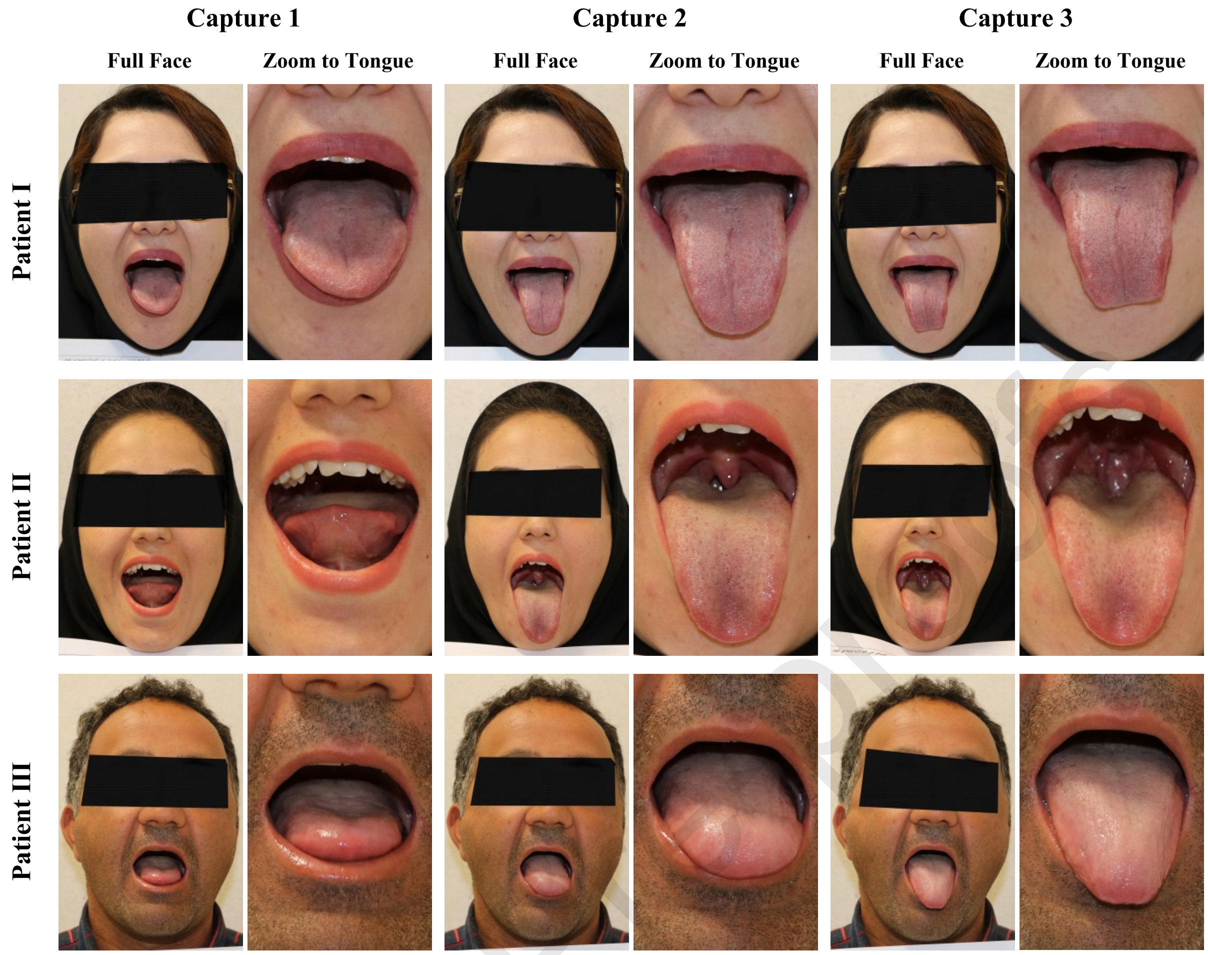


Figure 1: Example tongue photos in our dataset.

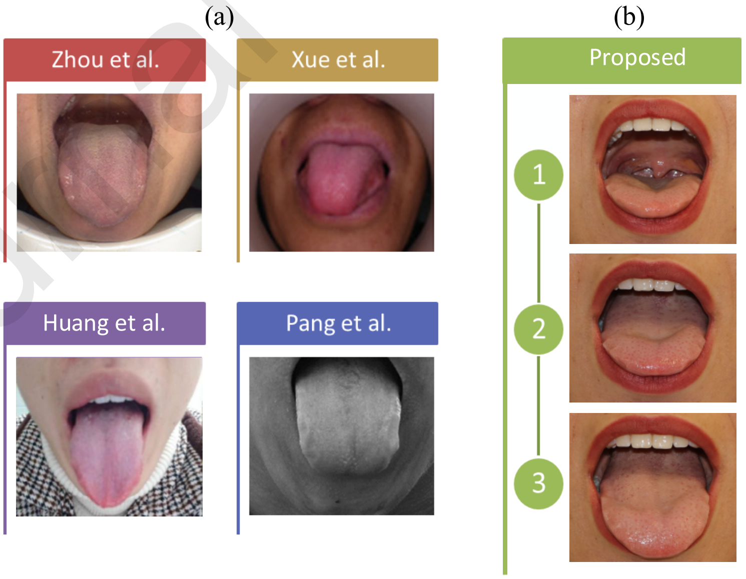


Figure 2: Comparison of human tongue imaging. (a) Only a single photo was taken per subject in existing tongue datasets. (b) Three tongue photos are captured in succession for each subject in the proposed dataset.

2.2. U-Net

U-Net [56] is a widely used deep neural network for medical image segmentation, which consists of an encoding path, a decoding path, and the feature map concatenation procedure. Let$X^{i,j}$ denotes a node in the U-Net architecture, where i represents the depth, and j represents either encoder (i.e., $j = 0$) or decoder (i.e., $j = 1$) side of the U-Net architecture. The encoder (i.e., $X^{i,0}$; $i = [0, 1, 2, 3, 4]$) of the U-Net extracts high-level features from input images by using convolution and pooling layers (see Figure 3). On the other hand, the decoding path of the U-Net (i.e., $X^{i,1}$; $i = [0, 1, 2, 3]$) expands the encoder-learned features sequentially using deconvolution and up-pooling layers and finally produces the segmentation mask. In addition, the encoding feature maps at $X^{i,0}$ is concatenated to the decoding feature map at $X^{i,1}$ via skip connections to recover the spatial information effectively.

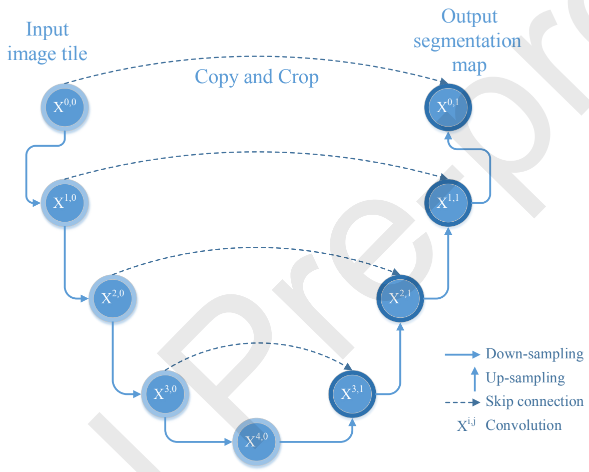


Figure 3: The U-Net architecture used in this work.

2.3. Learning-to-Augment Strategy

In this paper, we incorporated the learning-to-augment strategy [52, 57] to improve the robustness and generalizability of U-Net. The learning-to-augment strategy uses a noisy image generator, a controller, an augmenter, and a few child networks. This strategy automatically selects optimal parameters for the Gaussian, speckle, and Poisson noise to generate new augmented images. The mean (μ) and variance (v) parameters are used for the Gaussian and speckle noise [58-61]. Figure 4 shows a few photographic images corrupted by different noises. We see in Figure 5, the noisy image generator module adds noise to the original photographic images with given noise parameters. In the first iteration, parameters (μ, v) of the Gaussian, speckle, and Poisson noise are chosen randomly. From the second iteration onwards, the controller module updates these parameters by choosing the optimal policy. Then, the augmenter module generates new augmented images by adding noise to the original images. Afterward, child networks are trained using newly generated augmented images to assess the network performance on a given task. A child network is a U-Net-based model trained with an augmentation policy. Finally, the controller (a Bayesian optimizer [62, 63]) replaces weak policies (i.e., noise parameters) with a newer set of stronger augmentation policies by searching the parameter search space. Figure 6 shows how a child network is trained with the new data generated by an augmentation policy.

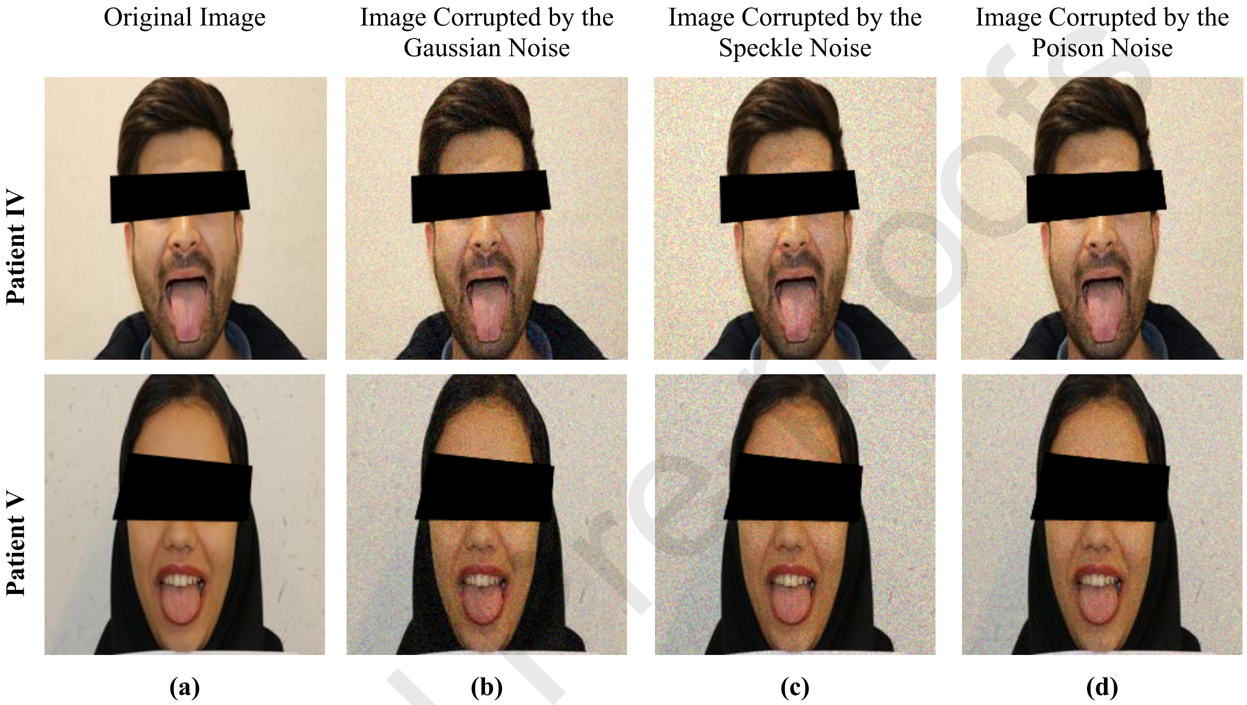

Original Image
Image Corrupted by the Gaussian Noise
Image Corrupted by the Speckle Noise
Image Corrupted by the Poison Noise
Patient IV
Patient V
(a)
(b)
(c)
(d)

Figure 4. Noisy photographic images corrupted by the Gaussian, speckle, and Poisson noises. (a) Original images, (b)-(d) images corrupted by the Gaussian ($\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{v} = \mathbf{0.003}$), speckle ($\boldsymbol{v} = \mathbf{0.01}$), and Poisson noises, respectively.

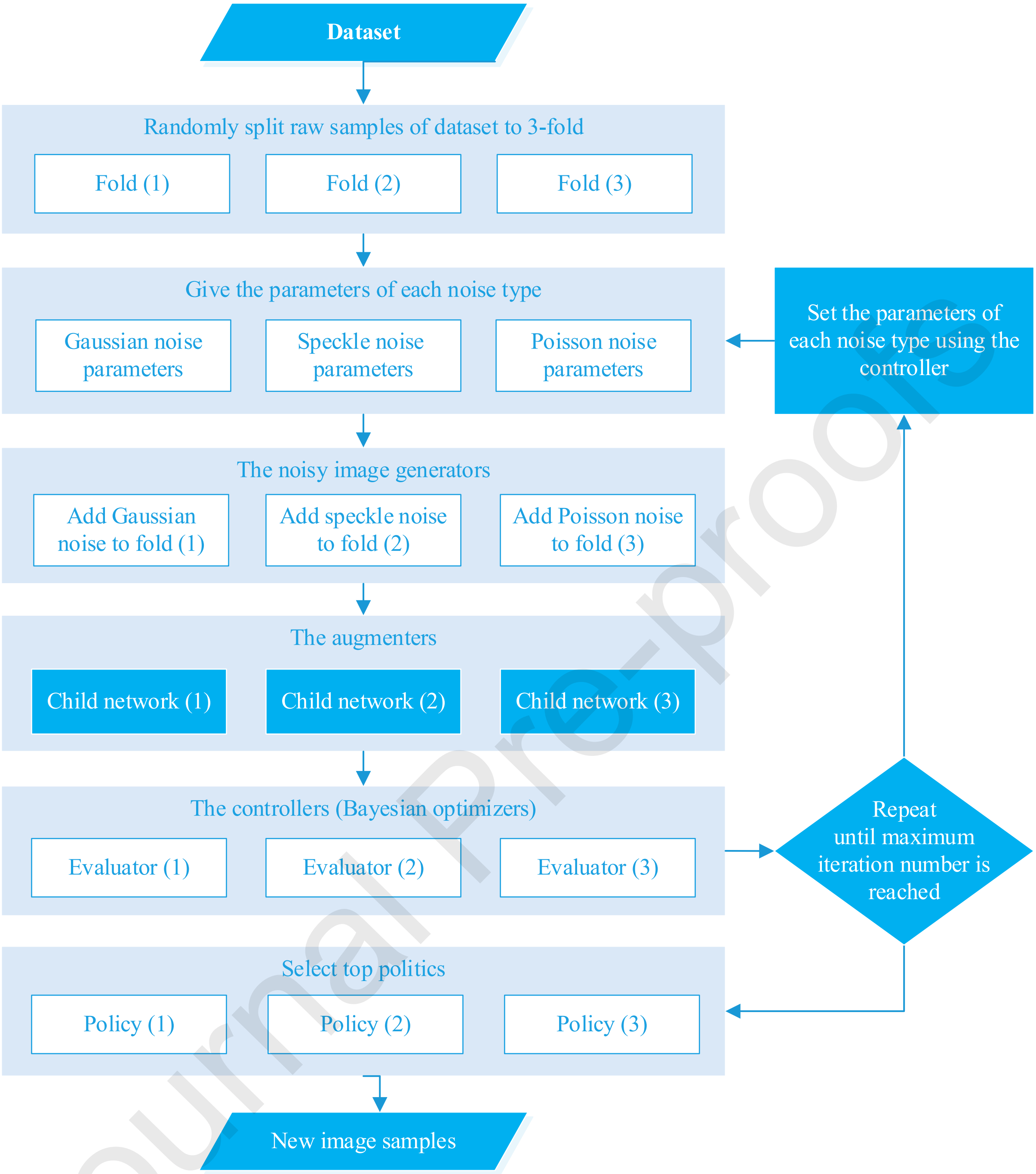
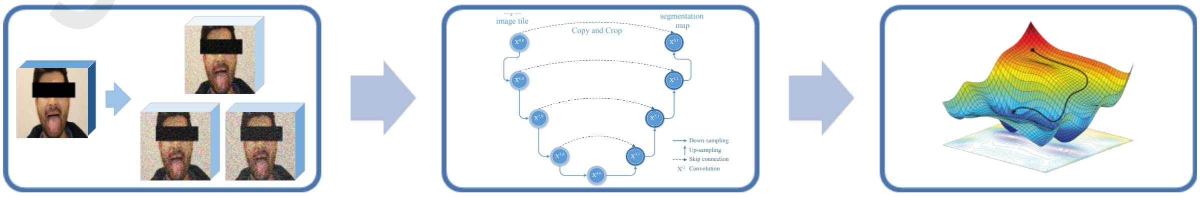

Dataset
Randomly split raw samples of dataset to 3-fold
Fold (1)
Fold (2)
Fold (3)
Give the parameters of each noise type
Gaussian noise parameters
Speckle noise parameters
Poisson noise parameters
Set the parameters of each noise type using the controller
The noisy image generators
Add Gaussian noise to fold (1)
Add speckle noise to fold (2)
Add Poisson noise to fold (3)
The augmenters
Child network (1)
Child network (2)
Child network (3)
The controllers (Bayesian optimizers)
Evaluator (1)
Evaluator (2)
Evaluator (3)
Repeat until maximum iteration number is reached
Select top politics
Policy (1)
Policy (2)
Policy (3)
New image samples

Figure 5. Flowchart showing our learning-to-augment strategy [51, 52].



(a) (b) (c)

Figure 6: The U-Net-based child network is trained with the new data generated by an augmentation policy. (a) Generation of noisy images by corrupting original images, (b) a child network to be trained on the new data, and (c) the controller finds the optimal noise parameters.

2.4. Proposed LAIU-Net

The proposed LAIU-Net combines the learning-to-augment strategy and the U-Net with a depth of 5 encoder stages (i.e., $X^{i,0}; i = [0, 1, 2, 3, 4]$) and 4 decoder stages (i.e., $X^{i,1}$; $i = [0, 1, 2, 3]$) (see Figure 7). As shown in Figure 7 and Table 2, the U-Net architecture of the proposed LAIU-Net contains 18 convolution layers, 4 transposed convolution, 22 leaky rectified linear unit (ReLU) activation layers, 4 pooling layers, 4 layers of batch normalization, and a Dropout layer. We incorporated the leaky ReLU to improve the performance of the proposed LAIU-Net optimization and segmentation [64]. We also incorporated a Dropout layer in the proposed LAIU-Net to avoid the overfitting problem. The dropout layer randomly drops $(1 - p)\%$ neurons while keeping $p\%$ neurons active in a particular layer. Finally, we used batch normalization and some conventional data augmentation methods for stable training of LAIU-Net. Although batch normalization optimizes network training, training may get slower because of the extra calculations of batch normalization during the forward pass. To reduce the computational overhead, we used 4 batch normalization layers at the end of encoder blocks at four different depths (see Figure 7).

3. Validation

3.1. Statistical Metrics

To evaluate the performance of the proposed LAIU-Net and state-of-the-art U-Net with different configurations on the tongue image segmentation, we use two quantitative metrics such as the intersection over union (IoU), and mean boundary F1 Score (BF Score) defined as:

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP}} \quad (2)$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad (3)$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (4)$$

$$\mathrm{BF\ Score} = \frac{2 \times \mathrm{Precision} \times \mathrm{Reall}}{\mathrm{Precision} + \mathrm{Reall}} \quad (5)$$

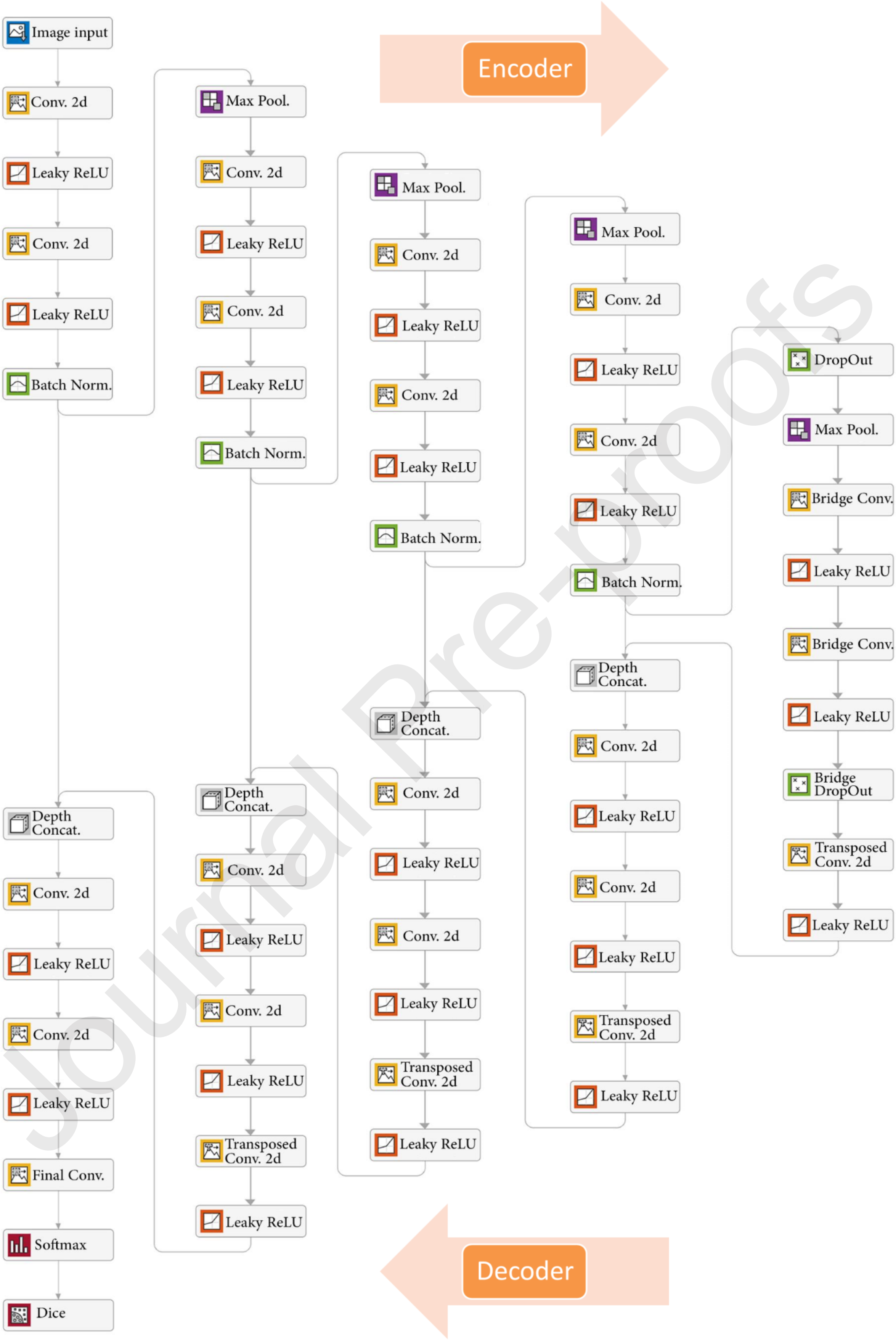


Figure 7: The architecture of the improved U-Net used in the proposed LAIU-Net

Table 2: Comparison of configurations of the improved U-Net (part of LAIU-Net) and 5 other U-Net architectures. Acronyms- Config: Configuration

#	Layers	Config. I	Config. II	Config. III	Config. IV	Config. V	LAIU-Net	Description
1	Image Input	Original Data	Original Data	Original Data	Original Data	Conventional Data Augmentation	Learning-to-Augment	256×256×3 images with 'zerocenter' normalization
2	Encoder-Stage-1	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	64 3×3 convolutions stride [1 1] and padding 'same'
3	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
4	Encoder-Stage-1	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	64 3×3 convolutions stride [1 1] and padding 'same'
5	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
6	BN	-	-	BN	BN	BN	BN	Batch normalization
7	Encoder-Stage-1	Pooling	Pooling	Pooling	Pooling	Pooling	Pooling	2×2 max pooling stride [2 2] and padding [0 0 0 0]
8	Encoder-Stage-2	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	128 3×3 convolutions stride [1 1] and padding 'same'
9	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
10	Encoder-Stage-2	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	128 3×3 convolutions stride [1 1] and padding 'same'
11	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
12	BN	-	-	BN	BN	BN	BN	Batch normalization
13	Encoder-Stage-2	Pooling	Pooling	Pooling	Pooling	Pooling	Pooling	2×2 max pooling stride [2 2] and padding [0 0 0 0]
14	Encoder-Stage-3	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	256 3×3 convolutions stride [1 1] and padding 'same'
15	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
16	Encoder-Stage-3	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	256 3×3 convolutions stride [1 1] and padding 'same'
17	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
18	BN	-	-	BN	BN	BN	BN	Batch normalization
19	Encoder-Stage-3	Pooling	Pooling	Pooling	Pooling	Pooling	Pooling	2×2 max pooling stride [2 2] and padding [0 0 0 0]
20	Encoder-Stage-4	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	512 3×3 convolutions stride [1 1] and padding 'same'
21	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
22	Encoder-Stage-4	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	512 3×3 convolutions stride [1 1] and padding 'same'
23	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
24	BN	-	-	BN	BN	BN	BN	Batch normalization
25	Encoder-Stage-4	Dropout	Dropout	Dropout	Dropout	Dropout	Dropout	50% dropout
26	Encoder-Stage-4	Pooling	Pooling	Pooling	Pooling	Pooling	Pooling	2×2 max pooling stride [2 2] and padding [0 0 0 0]
27	Bridge	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	1024 3×3 convolutions stride [1 1] and padding 'same'
28	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
29	Bridge	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	1024 3×3 convolutions stride [1 1] and padding 'same'
30	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
31	Bridge	Dropout	Dropout	Dropout	Dropout	Dropout	Dropout	50% dropout
32	Decoder-Stage-1	Transposed Conv.	Transposed Conv.	Transposed Conv.	Transposed Conv.	Transposed Conv.	Transposed Conv.	512 2×2 transposed convolutions stride [2 2] and cropping [0 0 0 0]

33	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
34	Decoder-Stage-1	Depth Concat.	Depth Concat.	Depth Concat.	Depth Concat.	Depth Concat.	Depth Concat.	Depth concatenation of 2 inputs
35	Decoder-Stage-1	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	512 3×3 convolutions stride [1 1] and padding 'same'
36	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
37	Decoder-Stage-1	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	512 3×3 convolutions stride [1 1] and padding 'same'
38	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
39	Decoder-Stage-2	Transposed Conv.	Transposed Conv.	Transposed Conv.	Transposed Conv.	Transposed Conv.	Transposed Conv.	256 2×2 transposed convolutions stride [2 2] and cropping [0 0 0 0]
40	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
41	Decoder-Stage-2	Depth Concat.	Depth Concat.	Depth Concat.	Depth Concat.	Depth Concat.	Depth Concat.	Depth concatenation of 2 inputs
42	Decoder-Stage-2	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	256 3×3 convolutions stride [1 1] and padding 'same'
43	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
44	Decoder-Stage-2	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	256 3×3 convolutions stride [1 1] and padding 'same'
45	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
46	Decoder-Stage-3	Transposed Conv.	Transposed Conv.	Transposed Conv.	Transposed Conv.	Transposed Conv.	Transposed Conv.	128 2×2 transposed convolutions stride [2 2] and cropping [0 0 0 0]
47	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
48	Decoder-Stage-3	Depth Concat.	Depth Concat.	Depth Concat.	Depth Concat.	Depth Concat.	Depth Concat.	Depth concatenation of 2 inputs
49	Decoder-Stage-3	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	128 3×3 convolutions stride [1 1] and padding 'same'
50	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
51	Decoder-Stage-3	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	128 3×3 convolutions stride [1 1] and padding 'same'
52	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
53	Decoder-Stage-4	Transposed Conv.	Transposed Conv.	Transposed Conv.	Transposed Conv.	Transposed Conv.	Transposed Conv.	64 2×2 transposed convolutions stride [2 2] and cropping [0 0 0 0]
54	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
55	Decoder-Stage-4	Depth Concat.	Depth Concat.	Depth Concat.	Depth Concat.	Depth Concat.	Depth Concat.	Depth concatenation of 2 inputs
56	Decoder-Stage-4	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	64 3×3 convolutions stride [1 1] and padding 'same'
57	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
58	Decoder-Stage-4	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	64 3×3 convolutions stride [1 1] and padding 'same'
59	Activation Fun.	ReLU	ReLU	ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU with scale 0.01
60	Final-Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	Conv.	2 1×1 convolutions stride [1 1] and padding 'same'
61	Activation Fun.	Sigmoid	Softmax	Softmax	Softmax	Softmax	Softmax	Softmax
62	Dice		Dice Pixel Classification Layer					Dice Pixel Classification Layer

3.2. Implementation

We use the deep network designer of MATLAB 2021b (MathWorks Inc., Natick, MA) to train, validate, and test our proposed LAIU-Net in an Intel(R) Core(TM) i7-7700HQ CPU 2.81 GHz with 32 GB of RAM and 8 GB of VRAM. We resized our original image of size 6000 × 4000 × 3 pixels to 256 × 256 × 3 pixels before feeding it to our deep model. Then we employed "imresize" function of MATLAB (it uses bicubic interpolation) to resize our original image to 256×256×3 pixels before feeding it to our deep model. We used the stochastic gradient descent with momentum (SGDM) as an optimizer with a maximum epoch of 5, minimum batch size of 5, a momentum of 0.9, an initial learning rate of 0.001, and dropout $p = 0.5$. We used approximately 80% of the photographic tongue data for training and validation and approximately 20% (68 images) of the data for testing. We made sure that images from a particular patient is not split between the training, validation, and test sets. We also show curves for train and validation accuracy *vs.* iteration for the network training duration for the task of tongue image segmentation in Figure 8.

4. Results and Discussion

We provide a comparative tongue image segmentation performance of our LAIU-Net and 5 different configurations of U-Net architectures. We can see in Table 2 that configuration I uses sigmoid activation. On the other hand, configuration II uses SoftMax activation. In configuration III, 4 batch normalization layers have been added to the U-Net. In configuration IV, leaky ReLU is used instead of ReLU. As conventional data augmentation methods, we performed a rotation of 90° and a vertical flip of images in configuration V. The proposed LAIU-Net incorporates a learning-to-augment strategy in addition to the mentioned data augmentation. The proposed LAIU-Net uses the best set of parameters for the Gaussian, speckle, and Poisson noises to be applied to original photographic tongue images for data augmentation. These best sets of noise parameters are chosen as the best policy by the learning-to-augment strategy.

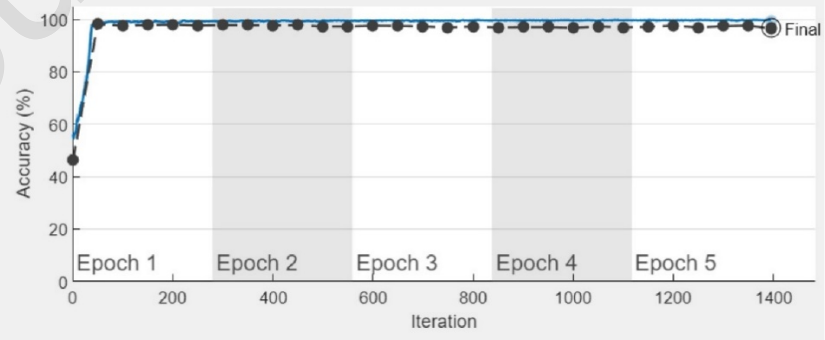


Figure 8. Accuracy curves of the LAIU-Net training and validation for the tongue image segmentation.

4.1. Quantitative Performance Comparison

At first, we show the confusion matrices of the tongue image segmentation mask by the proposed LAIU-Net and other U-Net configurations in Figure 9. In the matrix cells, we show the number of pixels count that fall inside either predicted tongue or background classes. We see in Figure 9 that the number of pixels representing the tongue body is much less than that of the background. So, there is a class imbalance in the dataset, and therefore, IoU and BF Score are the most suitable metrics to evaluate the performance of this study. We also plot the mean IoU and mean BF Score of the segmented tongue by different methods in Figure 10. Here, we see that the mean IoU (91.5%) and mean BF Score (93.1%) by the proposed LAIU-Net are the best compared to that by other techniques.

Predicted

Actual		**Tongue**	**Background**	
	Tongue	56149	14321	79.68%
	Background	19511	4366467	99.56%
		74.21%	99.67%	

U-Net with Configuration I

Predicted

Actual		**Tongue**	**Background**	
	Tongue	59011	11459	83.74%
	Background	21671	4364307	99.51%
		73.14%	99.74%	

U-Net with Configuration II

Predicted

Actual		**Tongue**	**Background**	
	Tongue	56256	14214	79.83%
	Background	10481	4375497	99.76%
		84.30%	99.68%	

U-Net with Configuration III

Predicted

Actual		**Tongue**	**Background**	
	Tongue	56639	13831	80.37%
	Background	10004	4375974	99.77%
		84.99%	99.68%	

U-Net with Configuration IV

Predicted

Actual		**Tongue**	**Background**	
	Tongue	61116	9354	86.73%
	Background	9260	4376718	99.79%
		86.84%	99.79%	

U-Net with Configuration V

Predicted

Actual		**Tongue**	**Background**	
	Tongue	65476	4994	92.91%
	Background	8059	4377919	99.82%
		89.04%	99.89%	

The proposed LAIU-Net

Figure 9. Confusion matrices showing the tongue image segmentation performance by different methods.

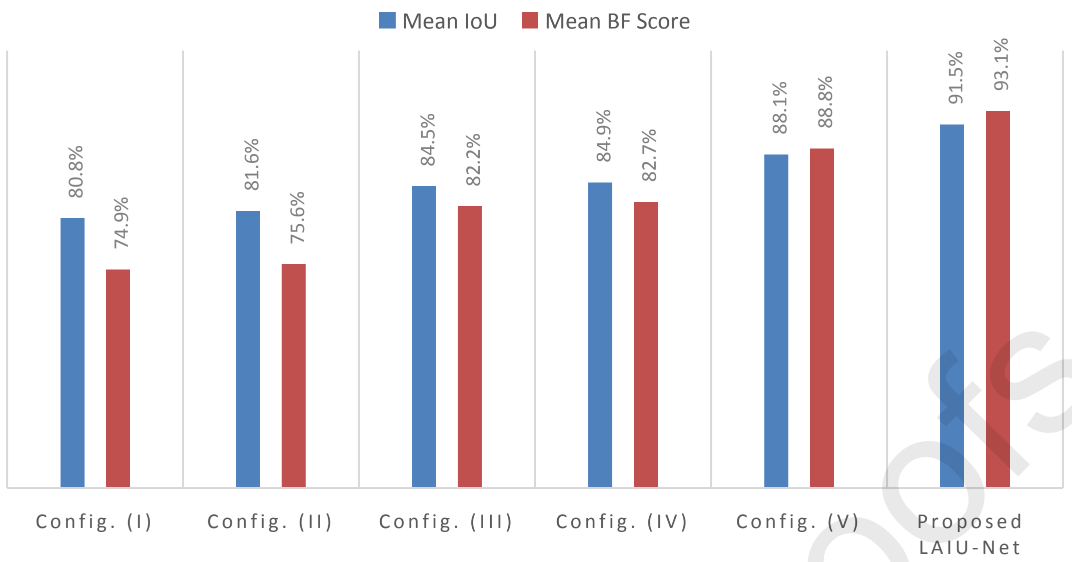


Figure 10. Tongue semantic segmentation performance by different methods. Acronym- Config: Configuration.

We show the semantic segmentation performance of different methods for two positive classes (i.e., tongue and background) in terms of IoU and mean BF Score in Table 3. At first, we considered "tongue" as the positive class and "background" as the negative class. We see in Table 3 that the proposed LAIU-Net performed the best in tongue segmentation in terms of IoU (83.38%) and mean BF Score (87.80%) compared to other configurations of U-Net. The closest performance is shown by the U-Net with configuration IV. Similarly, when "background" is considered as the positive class, the proposed LAIU-Net showed the best performance in terms of IoU (99.70%) and mean BF Score (98.37%).

Table 3. Evaluating the performance of segmentation with deferent positive classes

Method	**Positive class**	**IoU (%)**	**Mean BF Score (%)**
U-Net with Configuration I	Tongue	0.6240	0.5727
	Background	0.9923	0.9262
U-Net with Configuration II	Tongue	0.6404	0.5936
	Background	0.9925	0.9176
U-Net with Configuration III	Tongue	0.6949	0.6911
	Background	0.9944	0.9526
U-Net with Configuration IV	Tongue	0.7038	0.7009
	Background	0.9946	0.9529
U-Net with Configuration V	Tongue	0.7665	0.8056
	Background	0.9958	0.9701
Proposed LAIU-Net	Tongue	0.8338	0.8780
	Background	0.9970	0.9837

In Figure 11, we show mean IoU by different methods on male (26 images) and female (42 images) patients separately to see the effect of gender on the semantic segmentation performance. As we see in Figure 11 that there is no significant difference in segmentation performance on the basis of gender.

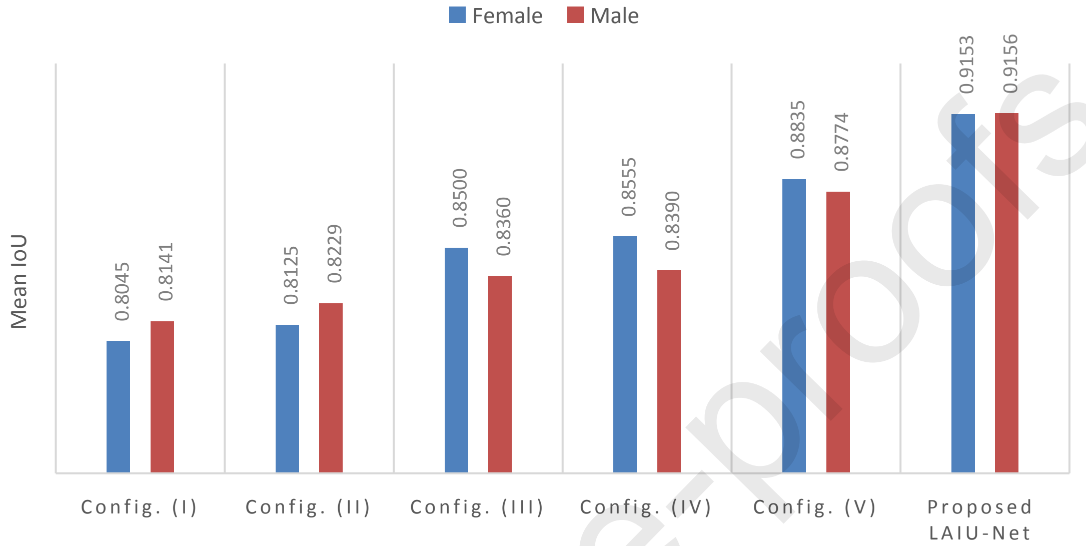


Figure 11. The semantic segmentation performance of different methods separately on male (26 images) and female (42 images) in terms of mean IoU.

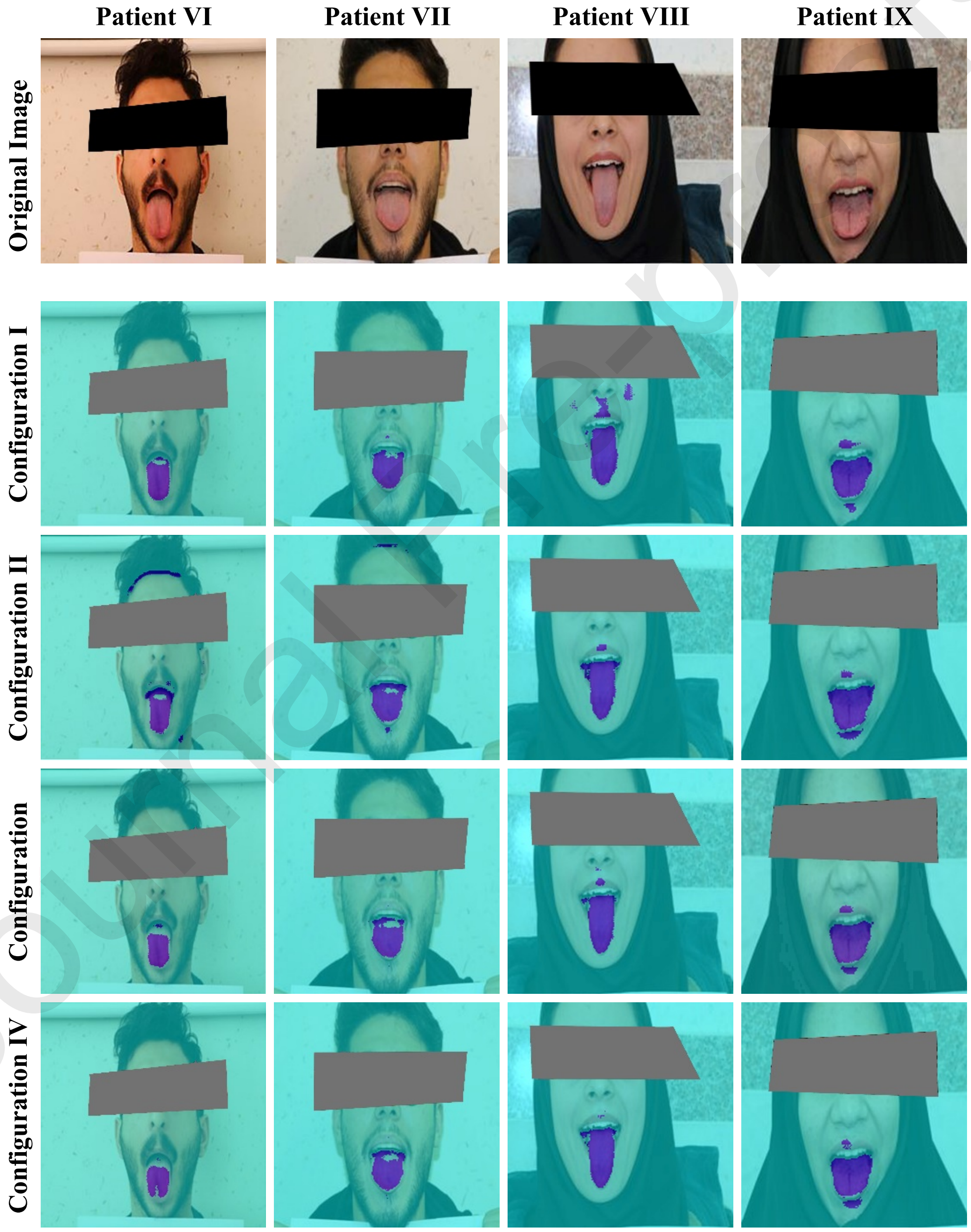
4.2. Qualitative Performance Comparison

Finally, we show the qualitative tongue segmentation performance by the proposed LAIU-Net and other U-Net-based techniques in

Figure **12**. We see in this figure that the tongue masks produced by the proposed LAIU-Net are the best compared to that by other U-Net configurations.

5. Conclusion

In this paper, we proposed the first learning-to-augment strategy-incorporated robust and generalized U-Net architecture for photographic tongue image segmentation. We produced training optimal augmented images via a learning-to-augment strategy that uses the Bayesian optimizer to choose the best set of policies for data augmentation via selecting the optimized Gaussian, speckle, and Poisson noise parameters. We showed the efficacy of our proposed learning-to-augment strategy-incorporated U-Net on the depressed human's tongue segmentation task and demonstrated the best performance in terms of IoU and mean BF Score compared to other U-Net configurations. In addition, we built a photographic tongue dataset consisting of 333 human tongue images of 111 patients diagnosed with depression. Despite promising results, our study has several limitations. For example, our dataset is small which hinders us to study the effect of sex on tongue segmentation performance. In addition, the age distribution of our patients is not uniform which further hinders us to study the effect of age on tongue segmentation performance. Therefore, in the future, we aim to extend this work with additional experiments after collecting more tongue images from depressed as well as healthy human subjects.



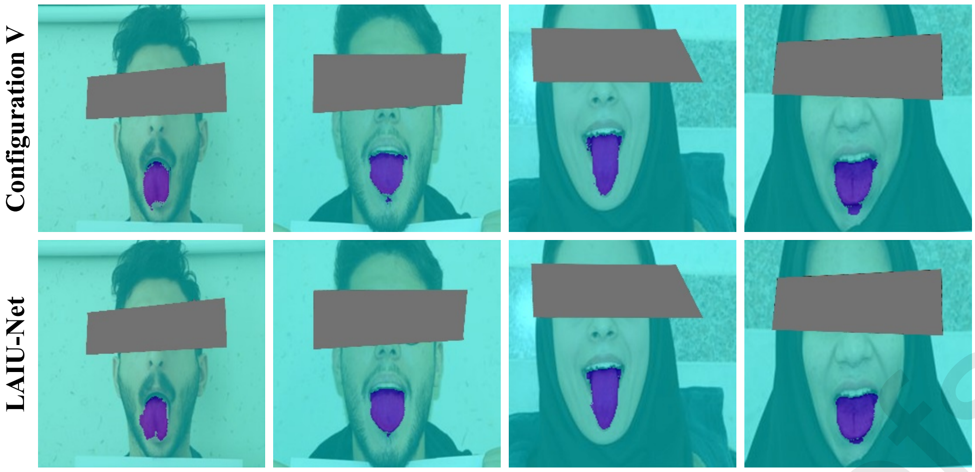


Figure 12. Semantic tongue segmentation performance by different methods.

References

[1] C.J. Jung, Y.J. Jeon, J.Y. Kim, K.H. Kim, Review on the current trends in tongue diagnosis systems, Integrative Medicine Research, 1 (2012) 13-20.

[2] J. Hernandez, C. Ferguson, A. Sano, W. Chen, W. Li, A.S. Yeung, R.W. Picard, Stress measurement from tongue color imaging, 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), 2017, pp. 152-157.

[3] C. Chang, C. Lin, Development of a TCM-based pulse impedance measurement system, 2017 IEEE/SICE International Symposium on System Integration (SII), 2017, pp. 499-504.

[4] M.-C. Hu, K.-C. Lan, W.-C. Fang, Y.-C. Huang, T.-J. Ho, C.-P. Lin, M.-H. Yeh, P. Raknim, Y.-H. Lin, M.-H. Cheng, Y.-T. He, K.-C. Tseng, Automated tongue diagnosis on the smartphone and its applications, Computer Methods and Programs in Biomedicine, 174 (2019) 51-64.

[5] M.H. Tania, K. Lwin, M.A. Hossain, Advances in automated tongue diagnosis techniques, Integrative Medicine Research, 8 (2019) 42-56.

[6] S.-Y. Kim, S.H. Hong, J.-W. Park, H. Lee, J. Kim, Y. Kim, Y.-S. Baik, S.-J. Ko, S.-K. Kim, I.-S. Lee, Y. Chae, H.-J. Park, Analysis of acupuncture diagnostic decision from the clinical information of a functional dyspepsia patient, Integrative Medicine Research, 9 (2020) 100419.

[7] Q. Zhang, J. Zhou, B. Zhang, Computational Traditional Chinese Medicine diagnosis: A literature survey, Computers in Biology and Medicine, 133 (2021) 104358.

[8] J. Ma, G. Wen, C. Wang, L. Jiang, Complexity perception classification method for tongue constitution recognition, Artificial Intelligence in Medicine, 96 (2019) 123-133.

[9] G. Arji, R. Safdari, H. Rezaeizadeh, A. Abbassian, M. Mokhtaran, M. Hossein Ayati, A systematic literature review and classification of knowledge discovery in traditional medicine, Computer Methods and Programs in Biomedicine, 168 (2019) 39-57.

[10] C.-C. Chiu, A novel approach based on computerized image analysis for traditional Chinese medical diagnosis of the tongue, Computer Methods and Programs in Biomedicine, 61 (2000) 77-89.

[11] H.Z. Zhang, K.Q. Wang, D. Zhang, B. Pang, B. Huang, Computer Aided Tongue Diagnosis System, 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, 2005, pp. 6754-6757.

[12] P. Bo, D. Zhang, W. Kuanquan, The bi-elliptical deformable contour and its application to automated tongue segmentation in Chinese medicine, IEEE Transactions on Medical Imaging, 24 (2005) 946-956.

[13] K. Wu, D. Zhang, Robust tongue segmentation by fusing region-based and edge-based approaches, Expert Systems with Applications, 42 (2015) 8027-8038.

[14] Z. Wangmeng, W. Kuanquan, D. Zhang, Z. Hongshi, Combination of polar edge detection and active contour model for automated tongue segmentation, Third International Conference on Image and Graphics (ICIG'04), 2004, pp. 270-273.

[15] S. Yu, J. Yang, Y. Wang, Y. Zhang, Color Active Contour Models Based Tongue Segmentation in Traditional Chinese Medicine, 2007 1st International Conference on Bioinformatics and Biomedical Engineering, 2007, pp. 1065-1068.

[16] Saparudin, Erwin, M. Fachrurrozi, Tongue Segmentation Using Active Contour Model, IOP Conference Series: Materials Science and Engineering, 190 (2017) 012041.

[17] M. Shi, G. Li, F. Li, C2G2FSnake: automatic tongue image segmentation utilizing prior knowledge, Science China Information Sciences, 56 (2013) 1-14.

[18] H. Zhang, W. Zuo, K. Wang, D. Zhang, A snake-based approach to automated segmentation of tongue image using polar edge detector, International Journal of Imaging Systems and Technology, 16 (2006) 103-112.

[19] An automatic tongue detection and segmentation framework for computer–aided tongue image analysis, International Journal of Functional Informatics and Personalised Medicine, 4 (2012) 56-68.

[20] J. Ning, D. Zhang, C. Wu, F. Yue, Automatic tongue image segmentation based on gradient vector flow and region merging, Neural Computing and Applications, 21 (2012) 1819-1826.

[21] C. Li, W. Dongyi, L. Yiqin, G. Xiaohang, S. Huiliang, A novel automatic tongue image segmentation algorithm: Color enhancement method based on L a b color space, 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2015, pp. 990-993.

[22] J.q. Du, Y.s. Lu, M.f. Zhu, K. Zhang, C.h. Ding, A Novel Algorithm of Color Tongue Image Segmentation Based on HSI, 2008 International Conference on BioMedical Engineering and Informatics, 2008, pp. 733-737.

[23] M. Shi, G. Li, F. Li, C. Xu, A novel tongue segmentation approach utilizing double geodesic flow, 2012 7th International Conference on Computer Science & Education (ICCSE), 2012, pp. 21-25.

[24] M.-J. Shi, G.-Z. Li, F.-F. Li, C. Xu, Computerized tongue image segmentation via the double geo-vector flow, Chinese Medicine, 9 (2014) 7.

[25] C. Liang, D. Shi, A Prior Knowledge-Based Algorithm for Tongue Body Segmentation, 2012 International Conference on Computer Science and Electronics Engineering, 2012, pp. 646-649.

[26] Z. Cui, W. Zuo, H. Zhang, D. Zhang, Automated Tongue Segmentation Based on 2D Gabor Filters and Fast Marching, in: C. Sun, F. Fang, Z.-H. Zhou, W. Yang, Z.-Y. Liu (Eds.) Intelligence Science and Big Data Engineering, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 328-335.

[27] J. Guo, Y. Yang, Q. Wu, J. Su, F. Ma, Adaptive active contour model based automatic tongue image segmentation, 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2016, pp. 1386-1390.

[28] M.A. Hussain, G. Hamarneh, T.W. O'Connell, M.F. Mohammed, R. Abugharbieh, Segmentation-Free Estimation of Kidney Volumes in CT with Dual Regression Forests, in: L. Wang, E. Adeli, Q. Wang, Y. Shi, H.-I. Suk (Eds.) Machine Learning in Medical Imaging, Springer International Publishing, Cham, 2016, pp. 156-163.

[29] Z. Liu, J.-q. Yan, D. Zhang, Q.-L. Li, Automated tongue segmentation in hyperspectral images for medicine, Appl. Opt., 46 (2007) 8328-8334.

[30] F. Zhicheng, L. Wei, L. Xiaoqiang, L. Fufeng, W. Yiqin, Automatic tongue location and segmentation, 2008 International Conference on Audio, Language and Image Processing, 2008, pp. 1050-1055.

[31] W. Yuan, C. Liu, Cascaded CNN for Real-time Tongue Segmentation Based on Key Points Localization, 2019 IEEE 4th International Conference on Big Data Analytics (ICBDA), 2019, pp. 303-307.

[32] B. Lin, J. Xie, C. Li, Y. Qu, Deeptongue: Tongue Segmentation Via Resnet, 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 1035-1039.

[33] P. Qu, H. Zhang, L. Zhuo, J. Zhang, G. Chen, Automatic Tongue Image Segmentation for Traditional Chinese Medicine Using Deep Neural Network, in: D.-S. Huang, V. Bevilacqua, P. Premaratne, P. Gupta (Eds.) Intelligent Computing Theories and Application, Springer International Publishing, Cham, 2017, pp. 247-259.

[34] Y. Xue, X. Li, P. Wu, J. Li, L. Wang, W. Tong, Automated Tongue Segmentation in Chinese Medicine Based on Deep Learning, in: L. Cheng, A.C.S. Leung, S. Ozawa (Eds.) Neural Information Processing, Springer International Publishing, Cham, 2018, pp. 542-553.

[35] X. Huang, H. Zhang, L. Zhuo, X. Li, J. Zhang, TISNet-Enhanced Fully Convolutional Network with Encoder-Decoder Structure for Tongue Image Segmentation in Traditional Chinese Medicine, Computational and Mathematical Methods in Medicine, 2020 (2020) 6029258.

[36] J. Zhou, Q. Zhang, B. Zhang, X. Chen, TongueNet: A Precise and Fast Tongue Segmentation System Using U-Net with a Morphological Processing Layer, Applied Sciences, 9 (2019) 3128.

[37] Z. Huang, J. Miao, H. Song, S. Yang, Y. Zhong, Q. Xu, Y. Tan, C. Wen, J. Guo, A novel tongue segmentation method based on improved U-Net, Neurocomputing, 500 (2022) 73-89.

[38] Q. Ruan, Q. Wu, J. Yao, Y. Wang, H.-W. Tseng, Z. Zhang, An Efficient Tongue Segmentation Model Based on U-Net Framework, International Journal of Pattern Recognition and Artificial Intelligence, 35 (2021) 2154035.

[39] L. Li, Z. Luo, M. Zhang, Y. Cai, C. Li, S. Li, An iterative transfer learning framework for cross-domain tongue segmentation, Concurrency and Computation: Practice and Experience, 32 (2020) e5714.

[40] C. Zhou, H. Fan, Z. Li, Tonguenet: Accurate Localization and Segmentation for Tongue Images Using Deep Neural Networks, IEEE Access, 7 (2019) 148779-148789.

[41] W. Liu, C. Zhou, Z. Li, Z. Hu, Patch-Driven Tongue Image Segmentation Using Sparse Representation, IEEE Access, 8 (2020) 41372-41383.

[42] R. Azadnia, A. Jahanbakhshi, S. Rashidi, M. khajehzadeh, P. Bazyar, Developing an automated monitoring system for fast and accurate prediction of soil texture using an image-based deep learning network and machine vision system, Measurement, 190 (2022) 110669.

[43] A. Jahanbakhshi, Y. Abbaspour-Gilandeh, K. Heidarbeigi, M. Momeny, Detection of fraud in ginger powder using an automatic sorting system based on image processing technique and deep learning, Computers in Biology and Medicine, 136 (2021) 104764.

[44] A. Jahanbakhshi, Y. Abbaspour-Gilandeh, K. Heidarbeigi, M. Momeny, A novel method based on machine vision system and deep learning to detect fraud in turmeric powder, Computers in Biology and Medicine, 136 (2021) 104728.

[45] A. Jahanbakhshi, M. Momeny, M. Mahmoudi, P. Radeva, Waste management using an automatic sorting system for carrot fruit based on image processing technique and improved deep neural networks, Energy Reports, 7 (2021) 5248-5256.

[46] M. Momeny, A. Jahanbakhshi, K. Jafarnezhad, Y.-D. Zhang, Accurate classification of cherry fruit using deep CNN based on hybrid pooling approach, Postharvest Biology and Technology, 166 (2020) 111204.

[47] M. Momeny, A. Jahanbakhshi, A.A. Neshat, R. Hadipour-Rokni, Y.-D. Zhang, Y. Ampatzidis, Detection of citrus black spot disease and ripeness level in orange fruit using learning-to-augment incorporated deep networks, Ecological Informatics, 71 (2022) 101829.

[48] M.A. Hussain, Z. Mirikharaji, M. Momeny, M. Marhamati, A.A. Neshat, R. Garbi, G. Hamarneh, Active deep learning from a noisy teacher for semi-supervised 3D image segmentation: Application to COVID-19 pneumonia infection in CT, Computerized Medical Imaging and Graphics, 102 (2022) 102127.

[49] M. Momeny, A.A. Neshat, A. Gholizadeh, A. Jafarnezhad, E. Rahmanzadeh, M. Marhamati, B. Moradi, A. Ghafoorifar, Y.-D. Zhang, Greedy Autoaugment for classification of mycobacterium tuberculosis image via generalized deep CNN using mixed pooling based on minimum square rough entropy, Computers in Biology and Medicine, 141 (2022) 105175.

[50] A. Jahanbakhshi, M. Momeny, M. Mahmoudi, Y.-D. Zhang, Classification of sour lemons based on apparent defects using stochastic pooling mechanism in deep convolutional neural networks, Scientia Horticulturae, 263 (2020) 109133.

[51] A. Akbarimajd, A.A. Neshat, M.A. Hussain, M. Momeny, Detection of Covid-19 in Noisy X-Ray Images Using Learning-to-Augment Incorporated Noise-Robust Deep CNN, Available at SSRN: https://ssrn.com/abstract=3979334.

[52] M. Momeny, A.A. Neshat, M.A. Hussain, S. Kia, M. Marhamati, A. Jahanbakhshi, G. Hamarneh, Learning-to-augment strategy using noisy and denoised data: Improving generalizability of deep

CNN for the detection of COVID-19 in X-ray images, Computers in Biology and Medicine, 136 (2021) 104704.

[53] J.B. Travers, Motor control of feeding and drinking, (2009).

[54] B. Pang, D. Zhang, K. Wang, The bi-elliptical deformable contour and its application to automated tongue segmentation in Chinese medicine, IEEE transactions on medical imaging, 24 (2005) 946-956.

[55] pp. Research Ethics Committees Certificate.

[56] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, in: N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi (Eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Cham, 2015, pp. 234-241.

[57] A. Akbarimajd, N. Hoertel, M.A. Hussain, A.A. Neshat, M. Marhamati, M. Bakhtoor, M. Momeny, Learning-to-augment incorporated noise-robust deep CNN for detection of COVID-19 in noisy X-ray images, Journal of Computational Science, 63 (2022) 101763.

[58] Y. Zhu, W. Shen, F. Cheng, C. Jin, G. Cao, Removal of high density Gaussian noise in compressed sensing MRI reconstruction through modified total variation image denoising method, Heliyon, 6 (2020) e03680.

[59] N. Karimi, M.R. Taban, A convex variational method for super resolution of SAR image with speckle noise, Signal Processing: Image Communication, 90 (2021) 116061.

[60] M. Momeny, A.M. Latif, M. Agha Sarram, R. Sheikhpour, Y.D. Zhang, A noise robust convolutional neural network for image classification, Results in Engineering, 10 (2021) 100225.

[61] M. Nooshyar, M. Momeny, Removal of high density impulse noise using a novel decision based adaptive weighted and trimmed median filter, 2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP), IEEE, 2013, pp. 387-391.

[62] P.I. Frazier, A tutorial on Bayesian optimization, arXiv preprint arXiv:1807.02811, (2018).

[63] M. Pelikan, D.E. Goldberg, E. Cantú-Paz, BOA: The Bayesian optimization algorithm, Proceedings of the genetic and evolutionary computation conference GECCO-99, Citeseer, 1999, pp. 525-532.

[64] A.L. Maas, A.Y. Hannun, A.Y. Ng, Rectifier nonlinearities improve neural network acoustic models, Proc. icml, Citeseer, 2013, pp. 3.

Highlights

1. We proposed the first learning-to-augment strategy-incorporated robust and generalized U-Net architecture for photographic tongue image segmentation.
2. We produced training optimal augmented images via selecting the optimized Gaussian, speckle, and Poisson noise parameters.
3. We showed the efficacy of our proposed learning-to-augment strategy-incorporated U-Net on the depressed human's tongue segmentation task.

Author Statement

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript.

Sincerely ,

Mahmoud Marhamati, Ali Asghar Latifi Zadeh, Mohammad Arafat Hussain, Kalegh Jafarnezhad, Ahad Jafarnezhad, Masoud Mojdehi far, Mehdi Bakhtoor, Mohammad Momeny

Conflict of Interest

The authors have declared no conflict of interest.

Sincerely ,

Mahmoud Marhamati, Ali Asghar Latifi Zadeh, Mohammad Arafat Hussain, Kalegh Jafarnezhad, Ahad Jafarnezhad, Masoud Mojdehi far, Mehdi Bakhtoor, Mohammad Momeny