Received: 7 February 2022 | Revised: 25 August 2022 | Accepted: 25 August 2022

DOI: 10.1002/mp.15969

RESEARCH ARTICLE

MEDICAL PHYSICS

A novel adaptive cubic quasi-Newton optimizer for deep learning based medical image analysis tasks, validated on detection of COVID-19 and segmentation for COVID-19 lung infection, liver tumor, and optic disc/cup

Yan Liu | **Maojun Zhang** | **Zhiwei Zhong** | **Xiangrong Zeng**

College of Systems Engineering, National University of Defense Technology, Changsha, China

Correspondence
Xiangrong Zeng, College of Systems Engineering, National University of Defense Technology, Changsha, 410073, China.
Email: zengxrong@foxmail.com

Funding information
Natural Science Foundation of Hunan Province, Grant/Award Number: 2019JJ50746; National Natural Science Foundation of China, Grant/Award Number: 61602494

Abstract

Background: Most of existing deep learning research in medical image analysis is focused on networks with stronger performance. These networks have achieved success, while their architectures are complex and even contain massive parameters ranging from thousands to millions in numbers. The nature of high dimension and nonconvex makes it easy to train a suboptimal model through the popular stochastic first-order optimizers, which only use gradient information.

Purpose: Our purpose is to design an adaptive cubic quasi-Newton optimizer, which could help to escape from suboptimal solution and improve the performance of deep neural networks on four medical image analysis tasks including: detection of COVID-19, COVID-19 lung infection segmentation, liver tumor segmentation, optic disc/cup segmentation.

Methods: In this work, we introduce a novel adaptive cubic quasi-Newton optimizer with high-order moment (termed ACQN-H) for medical image analysis. The optimizer dynamically captures the curvature of the loss function by diagonally approximated Hessian and the norm of difference between previous two estimates, which helps to escape from saddle points more efficiently. In addition, to reduce the variance introduced by the stochastic nature of the problem, ACQN-H hires high-order moment through exponential moving average on iteratively calculated approximated Hessian matrix. Extensive experiments are performed to access the performance of ACQN-H. These include detection of COVID-19 using COVID-Net on dataset COVID-chestxray, which contains 16 565 training samples and 1841 test samples; COVID-19 lung infection segmentation using Inf-Net on COVID-CT, which contains 45, 5, and 5 computer tomography (CT) images for training, validation, and testing, respectively; liver tumor segmentation using ResUNet on LiTS2017, which consists of 50 622 abdominal scan images for training and 26 608 images for testing; optic disc/cup segmentation using MRNet on RIGA, which has 655 color fundus images for training and 95 for testing. The results are compared with commonly used stochastic first-order optimizers such as Adam, SGD, and AdaBound, and recently proposed stochastic quasi-Newton optimizer Apollo. In task detection of COVID-19, we use classification accuracy as the evaluation metric. For the other three medical image segmentation tasks, seven commonly used evaluation metrics are utilized, that is, Dice, structure measure, enhanced-alignment measure (EM), mean absolute error (MAE), intersection over union (IoU), true positive rate (TPR), and true negative rate.

Results: Experiments on four tasks show that ACQN-H achieves improvements over other stochastic optimizers: (1) comparing with AdaBound, ACQN-H achieves 0.49%, 0.11%, and 0.70% higher accuracy on the COVID-chestxray dataset using network COVID-Net with VGG16, ResNet50 and DenseNet121 as backbones, respectively; (2) ACQN-H has the best scores in terms of evaluation metrics Dice, TPR, EM, and MAE on COVID-CT dataset using network Inf-Net. Particularly, ACQN-H achieves 1.0% better Dice as compared to Apollo; (3) ACQN-H achieves the best results on LiTS2017 dataset using network ResUNet, and outperforms Adam in terms of Dice by 2.3%; (4) ACQN-H improves the performance of network MRNet on RIGA dataset, and achieves 0.5% and 1.0% better scores on cup segmentation for Dice and IoU, respectively, compared with SGD. We also present fivefold validation results of four tasks. It can be found that the results on detection of COVID-19, liver tumor segmentation and optic disc/cup segmentation can achieve high performance with low variance. For COVID-19 lung infection segmentation, the variance on test set is much larger than on validation set, which may due to small size of dataset.
Conclusions: The proposed optimizer ACQN-H has been validated on four medical image analysis tasks including: detection of COVID-19 using COVID-Net on COVID-chestxray, COVID-19 lung infection segmentation using Inf-Net on COVID-CT, liver tumor segmentation using ResUNet on LiTS2017, optic disc/cup segmentation using MRNet on RIGA. Experiments show that ACQN-H can achieve some performance improvement. Moreover, the work is expected to boost the performance of existing deep learning networks in medical image analysis.

KEYWORDS
cubic quasi-Newton optimizer, high-order moment, medical image analysis

1 | INTRODUCTION

Deep learning for medical image analysis attracts great attention and achieves great success. It can take care of the simple repetitive and time-consuming process, help in diagnosing disease, evaluate prognosis, and plan operation.[1] Among numerous medical image analysis tasks, classification and segmentation are the most attractive.[2] Great efforts have been made on designing high-performance deep neural networks (DNNs), which can even exceed human recognition ability.[3] One of leading DNNs for medical image segmentation is UNet.[4] It adopts an encoder–decoder structure, which enables recovering full spatial resolution. Many variants of UNet[2,5–8] have been proposed for medical image analysis. For instance, H-DenseUNet[5] makes full use of adjacent computer tomography (CT) volume for model training and achieves improved performance on liver segmentation task. ResUNet[6] introduces a semantic segmentation model with context multiimages input and utilizes a new loss function that combines Dice loss[7] with cross-entropy loss, which brings faster convergence speed. MRNet[9] explores the utilization of rich annotation information from multiple experts and incorporates the multirater (dis-)agreement cues which help to generate better prediction.

Recently, COVID-19 ravages the world. The medical system suddenly suffers great pressure from exponentially increasing number of infections. This inspires the research on deep learning based COVID-19 automated diagnosis.[10] COVID-Net[11] is the first open-source convolutional neural network designed for COVID-19 detection and achieves good precision. COVID-Net and its many variants[12–14] augment the traditional healthcare strategy for tackling COVID-19, while they can hardly be applied to segment infected regions from CT slices faces. This may partly be due to high variation in infection characteristics, low-intensity contrast between infections and normal tissues, and lack of labeled data. To solve these problems, Inf-Net[15] is proposed to automatically identify infected regions from chest CT slices. It utilizes implicit reverse attention and explicit edge attention to improve the identification of infected regions.

Most of DNNs mentioned above are always complex with massive parameters ranging from thousands to millions in numbers. The high dimension and nonconvex nature of DNNs make them hard to optimize.[16,17] The most popular optimizers are the first-order ones that are based on first-order Taylor expansion of loss function. For many applications, SGD,[18] Adam,[19] and AdamW[20] are the default optimizers because of their simplicity and efficiency. Recently, due to its good

performance, Adam[19] also engenders an ever-growing list of modifications, such as AdaBound,[21] Radam,[22] Adabelief,[16] and Adax.[23] However, they are easy to be trapped in suboptimal solutions in algorithmic iterations, which largely due to only utilizing gradient information.[24] Instead, stochastic second-order optimizers can capture and exploit curvature properties of the loss landscape by incorporating both gradient and Hessian information, leading to better performance. AdaHessian,[24] as an adaptive Hessian-based optimizer, estimates the Hessian matrix diagonally, Hutchinson's method.[25] The method incorporates spatial averaging for Hessian diagonal which helps in denoising local Hessian information and enables AdaHessian to achieve better generalization. Instead of directly obtaining Hessian information, an alternative is a class of stochastic quasi-Newton optimizers,[26–28] which approximate the curvature of objective function only using gradient information. Generally, second-order optimizers require much more computational resource on both time and memory to calculate the Hessian matrix, while quasi-Newton optimizers are more applicable as they are able to balance performance and efficiency. As an example, for experiments on ImageNet using ResNext, the time cost of second-order optimizer AdaHessian can reach up to 11.78 and 9.58 times larger than those of first-order optimizer SGD and recent proposed stochastic quasi-Newton optimizer Apollo, respectively. Meanwhile, the memory cost of AdaHessian can reach up to 2.51 and 2.39 times larger than those of SGD and Apollo, respectively.[28]

In this work, we propose a novel adaptive cubic quasi-Newton optimizer with high-order moment (termed ACQN-H) for medical image analysis. Different from existing stochastic quasi-Newton optimizers which usually approximate Hessian only using the curvature of the loss function, the proposed optimizer incorporates both the curvature of the loss function and the norm of difference between previous two estimates. Besides, ACQN-H hires high-order moment through exponential moving average on iteratively calculated Hessian approximations to reduce the variance introduced by the stochastic nature of the problem. The performance of ACQN-H has been validated in four tasks including detection of COVID-19, COVID-19 lung infection segmentation, liver tumor segmentation, and optic disc/cup segmentation. Moreover, the work is also expected to boost the performance of other existing deep learning networks in medical image analysis.

Notation. We use italics letters such as ϵ and β to denote scalars, bold lowercase letters x and y to denote vectors, and bold uppercase letters H and D to denote matrices.

2 | METHODOLOGY

In this section, we first provide the formulation of the cubic quasi-Newton method in Subsection 2.1. Then, we describe the updated process for approximated Hessian matrix $\boldsymbol{B}_k$ in Subsection 2.2. Finally, the form of high-order moment applied for ACQN-H is discussed in Subsection 2.3.

2.1 | Formulation of cubic quasi-Newton method

Generally, the updated rule of the Newton method can be written as

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{H}_k^{-1}\boldsymbol{g}_k, \tag{1}$$

where $x_k \in \mathbb{R}^d$ is the parameter vector updated at kth iteration, $\boldsymbol{g}_k \in \mathbb{R}^d$ and $\boldsymbol{H}_k \in \mathbb{R}^{d\times d}$ are the gradient vector and Hessian matrix, respectively. Acquiring the exact Hessian $\boldsymbol{H}_k$ needs high computation cost. Instead, the quasi-Newton method approximates the second derivative of loss function as a series sum of first-order gradient information from prior iterations, and this is much more efficient.[28] The curvature of the loss function can be acquired through a second-order Tayler expansion

$$\boldsymbol{x}_{k+1} = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}_k) + \boldsymbol{g}_k^T(\boldsymbol{x}-\boldsymbol{x}_k) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_k)^T\boldsymbol{B}_k(\boldsymbol{x}-\boldsymbol{x}_k). \tag{2}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is the loss function and $\boldsymbol{x} \in \mathbb{R}^d$ is the weights of DNNs. The weight update process is shown as

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{B}_k^{-1}\boldsymbol{g}_k, \tag{3}$$

where arg min represents the set of values where the object function attains the minimum, $\boldsymbol{B_k} \in \mathbb{R}^{d\times d}$ is an approximation to the Hessian matrix at $\boldsymbol{x}_k$. To further enhance the global convergence, the cubic regularization is introduced[29] and the optimal weights can be acquired through finding the minimizer of cubically regularized second-order Taylor expansion

$$\boldsymbol{x}_{k+1} = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}_k) + \boldsymbol{g}_k^T(\boldsymbol{x}-\boldsymbol{x}_k) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_k)^T\boldsymbol{B}_k(\boldsymbol{x}-\boldsymbol{x}_k) + \frac{\rho}{6}||\boldsymbol{x}-\boldsymbol{x}_k||^3, \tag{4}$$

$\rho > 0$ is a sufficient large hyperparameter. By first-order optimality conditions, we set the derivative of the

objective to zero, which immediately yields

$$\boldsymbol{g}_k + \boldsymbol{B}_k(\boldsymbol{x} - \boldsymbol{x}_k) + \frac{\rho}{2}||\boldsymbol{x} - \boldsymbol{x}_k||(\boldsymbol{x} - \boldsymbol{x}_k) = 0, \quad (5)$$

which is a nonlinear system and can be approximated by a linear one as follows:

$$\boldsymbol{g}_k + \boldsymbol{B}_k(\boldsymbol{x} - \boldsymbol{x}_k) + \frac{\rho}{2}||\boldsymbol{x}_k - \boldsymbol{x}_{k-1}||(\boldsymbol{x} - \boldsymbol{x}_k) = 0 \quad (6)$$

yielding a novel update:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - (\boldsymbol{B}_k + \frac{\rho}{2}||\boldsymbol{x}_k - \boldsymbol{x}_{k-1}|| \cdot \boldsymbol{I})^{-1}\boldsymbol{g}_k, \quad (7)$$

where $\boldsymbol{I} \in \mathbb{R}^{d\times d}$ is the identity matrix. Comparing with (3), (7) additionally makes use of the norm of difference between previous two estimates, leading to better performance. Further, to guarantee the positive-definiteness, the Newton update in (7) is combined with rectifying operation and becomes

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{D}_k^{-1}\boldsymbol{g}_k, \quad (8)$$

where

$$\boldsymbol{D}_k = \max(\text{abs}(\boldsymbol{B}_k + \frac{\rho}{2}||\boldsymbol{x}_k - \boldsymbol{x}_{k-1}|| \cdot \boldsymbol{I}), \theta \cdot \boldsymbol{I}), \quad (9)$$

where θ is a positive parameter, $\max(\cdot,\cdot)$ operation enables $\boldsymbol{D}_k \in \mathbb{R}^{d\times d}$ to prevent the step size from becoming arbitrary large since there exists zero value in $\boldsymbol{B}_k$.

2.2 | Updating $\boldsymbol{B}_k$

For simplification, the matrix $\boldsymbol{B}_k$ in (8) is approximated by a diagonal matrix. Thus, acquiring $\boldsymbol{D}_k^{-1}$ at every iteration becomes computationally feasible since $\boldsymbol{D}_k$ is also a diagonal matrix. Here, $\boldsymbol{B}_k$ can be updated according to the quasi-Cauchy equation[30]:

$$\begin{aligned} &\boldsymbol{B}_{k+1} = \underset{\boldsymbol{B}}{arg\,\min}||\boldsymbol{B} - \boldsymbol{B}_k|| \\ &S.t. \boldsymbol{s}_k^T\boldsymbol{B}\boldsymbol{s}_k = \boldsymbol{s}_k^T\boldsymbol{y}_k (\text{weak scant equation}) \end{aligned} \quad (10)$$

where $\boldsymbol{s}_k = \boldsymbol{x}_k - \boldsymbol{x}_{k-1}$ and $\boldsymbol{y}_k = \boldsymbol{g}_k - \boldsymbol{g}_{k-1}$. The solution to the above problem with the Frobenius matrix based on the variational technique in Zhu et al.[30] is given by

$$\boldsymbol{B}_{k+1} = \boldsymbol{B}_k + \frac{\boldsymbol{s}_k^T\boldsymbol{y}_k - \boldsymbol{s}_k^T\boldsymbol{B}_k\boldsymbol{s}_k}{||\boldsymbol{s}_k||_4^4 + \epsilon}\text{Diag}(\boldsymbol{s}_k^2), \quad (11)$$

where $\text{Diag}(\boldsymbol{v})$ is the diagonal matrix with diagonal elements from vector $\boldsymbol{v}$.

2.3 | High-order moment

To reduce the variance and further improve the performance, we adapt the moments for both gradient and diagonally approximated Hessian. The first moment $\boldsymbol{m}_k \in \mathbb{R}^d$ is defined as

$$\boldsymbol{m}_k = \frac{(1-\beta_1^{k-1})\beta_1\boldsymbol{m}_{k-1} + (1-\beta_1)\boldsymbol{g}_k}{1-\beta_1^k}, \quad (12)$$

where β_1 is the first moment hyperparameter. The high-order moment $\boldsymbol{V}_k \in \mathbb{R}^{d\times d}$ is shown as follows:

$$\boldsymbol{V}_k = \frac{(1-\beta_2^{k-1})\beta_2\boldsymbol{V}_{k-1} + (1-\beta_2)\boldsymbol{D}_k^h}{1-\beta_2^k}, \quad (13)$$

where $\beta_2 \in (0, 1)$ is the hyperparameter and h represents the order of the second moment. The second moment utilizes historical second-order derivatives to smooth the noisy curvature information. Generally, in many image classification and segmentation tasks, Adam, AdamW, and AdaHessian would set $\beta_1 = 0.9$, $\beta_2 = 0.999$. We use the same setting to enable a fair comparison.

To summarize, the complete algorithm of adaptive cubic quasi-Newton optimizer with high-order moment (ACQN-H) is given in Algorithm 1. In which, at most first-order gradients are required, and $\boldsymbol{B}_k$, $\boldsymbol{D}_k$, and $\boldsymbol{V}_k$ are all diagonal. Therefore, ACQN-H iteratively updates with linear complexity for both time and memory.

2.4 | Performance evaluation

To access the performance of ACQN-H, the optimizer is extensively tested on a wide range of learning tasks: detection of COVID-19, COVID-19 lung infection segmentation, liver tumor segmentation, and optic disc/cup segmentation. The results in each task are compared with stochastic first-order optimizers like Adam,[19] SGD,[18] AdaBound,[21] and stochastic second-order optimizer Apollo.[28] Among them, Adam and SGD are the most common and default optimizers for these tasks, AdaBound is a recently proposed first-order optimizer that works well. For each task, fivefold cross-validation results are reported. The tested learning tasks are briefly explained below:

1. *Detection of COVID-19*: We experiment on the COVID-chestxray dataset[31] using COVID-Net.[27] The training set consists of 7966 normal chest X-rays and 8599 X-rays of negative samples, while the test set contains 885 normal X-rays and 956 negative X-rays. Besides, to broadly test its performance, we

ALGORITHM 1 ACQN-H

Require: n_g //Mini-batch size

Require: η //Stepsize

Require: $\beta_1, \beta_2 \in [0, 1)$ //Exponential decay rates for the moment estimates

Require: $\epsilon, \rho, \theta, h$ //Positive parameters

Require: $\boldsymbol{x}_0, \boldsymbol{g}_0, \boldsymbol{B}_0, \boldsymbol{m}_0, \boldsymbol{V}_0$ //Initialize variables

Require: $k \leftarrow 0$ //Initialize timestep

while $\boldsymbol{x}_k$ not converged **do**

$k \leftarrow k + 1$

$\boldsymbol{g}_k \leftarrow \nabla f_i(\boldsymbol{x}; \boldsymbol{\xi}_i)$ //Stochastic gradient at timestep k

$\boldsymbol{s}_k \leftarrow \boldsymbol{x}_k - \boldsymbol{x}_{k-1}$

$\boldsymbol{y}_k \leftarrow \boldsymbol{g}_k - \boldsymbol{g}_{k-1}$

$\boldsymbol{B}_k \leftarrow \boldsymbol{B}_{k-1} + \frac{\boldsymbol{s}_k^T \boldsymbol{y}_k - \boldsymbol{s}_k^T \boldsymbol{B}_{k-1} \boldsymbol{s}_k}{||\boldsymbol{s}_k||_4^4 + \epsilon} Diag(\boldsymbol{s}_k^2)$ //Update diagonal Hessian

$\boldsymbol{D}_k \leftarrow \max(abs(\boldsymbol{B}_k + \frac{\rho}{2}||\boldsymbol{x}_k - \boldsymbol{x}_{k-1}|| \cdot \boldsymbol{I}), \theta \cdot \boldsymbol{I})$ // Rectify for positive-definiteness

$\boldsymbol{m}_k \leftarrow \frac{(1-\beta_1^{k-1})\beta_1 \boldsymbol{m}_{k-1} + (1-\beta_1)\boldsymbol{g}_k}{1-\beta_1^k}$ //Update first moment

$\boldsymbol{V}_k \leftarrow \frac{(1-\beta_2^{k-1})\beta_2 \boldsymbol{V}_{k-1} + (1-\beta_2)\boldsymbol{D}_k^h}{1-\beta_2^k}$ //Update high-order second moment

$\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k - \eta \boldsymbol{V}_{\boldsymbol{k}}^{-\frac{1}{h}} \boldsymbol{m}_k$ //Update parameters

end while

return $\boldsymbol{x}_{k+1}$

set VGG16,[32] ResNet50,[33] and DenseNet121[34] as backbones instead of only one default ResNet50 backbone. Besides, the total training epoch is 300. Although there are image samples from the same patient, we only want to test the performance of the optimizer under the same conditions as those of COVID-Net.

2. *COVID-19 lung infection segmentation*: We report the performance of ACQN-H on the COVID-CT dataset[35] using a supervised version of the Inf-Net model.[15] COVID-CT is a COVID-19 CT segmentation dataset with 100 labeled CT slices, which consists of 45 CT images for training, 5 CT images for validation, and the remaining 50 images for testing. Additionally, the total training epoch is 100.
3. *Liver tumor segmentation*: We use ResUNet[6] on the LiTS2017 dataset,[36] which also served as a segmentation challenge during MICCAI 2017. The training set of LiTS2017 contains 50 622 abdominal scan images of 130 CT scans from 91 patients while the test set contains 26 608 images of 70 CT scans from 40 patients. The total training epoch is 300.
4. *Optic disc/cup segmentation*: We report experiments using the MRNet model[9] on the RIGA dataset,[37] which contains in total of 750 color fundus images. Followed with the experiment setting in MRNet, 655 samples are selected as the training set and 95 samples consist of the test set. Moreover, the total training epoch is 60.

Experiment environment. The deep neural network framework we experiment on is Pytorch1.7.1 with python3.6 and is GPU-accelerated. The hardware is a single RTX 3090Ti with I9-10920X CPU, while the RAM is 32GB.

Experiment setup. We perform a careful hyperparameter tuning in experiments as follows:

ACQN-H: We set $\theta = 1$, $\rho = 1$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$. As if a diagonal element of the approximated Hessian matrix $\boldsymbol{D}_k$ is less than 1, the corresponding element in D_k becomes 1. Thus, the update of this element can work as that of SGD and prevent the step to be arbitrarily large. Besides, we do not tune ρ and θ on different problems, which may help to reach a better result. The learning rate η is set to 0.1. In addition, for each task, we search the best order h from 2.0 to 10.0.

SGD: The momentum is set to 0.9, while the learning rate is searched among $\{a \times 10^b\}$, where $a \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and $b \in \{-4, -3, -2, -1\}$.

Adam, AdaBound, and Apollo: The learning rate is searched as SGD, and other parameters are set as their own default values in the literature.

Evaluation metrics. In the medical image classification task detection of COVID-19, the commonly used classification accuracy is utilized as the evaluation metric. For the other three medical image segmentation tasks, we integrate the default evaluation metrics into the following seven evaluation metrics, that is, Dice, structure measure (SM), enhanced-alignment measure (EM), mean absolute error (MAE), intersection over union (IoU), true positive rate (TPR), and true negative rate (TNR). Among these metrics, Dice, SM, EM, MAE, and IoU can measure the similarity between the result and ground truth. TPR means the correct rate of correctly segmented pixels of a target region, and TNR represents the correct proportion of background pixels that are segmented correctly. Assuming $\boldsymbol{T_0}$ and $\boldsymbol{T_1}$ represent the normal region and the ground truth (GT), respectively, $\boldsymbol{P_0}$ is the predicted normal region, and $\boldsymbol{P_1}$ means the predicted segmentation region, $|\cdot|$ means the number of matrix elements. The seven evaluation metrics can be formulated as follows:

1. Dice:

$$\boldsymbol{Dice}(\boldsymbol{T_1}, \boldsymbol{P_1}) = \frac{\boldsymbol{2}|\boldsymbol{T_1} \cap \boldsymbol{P_1}|}{|\boldsymbol{T_1}| \cup |\boldsymbol{P_1}|}. \quad (14)$$

2. Structure measure:

$$\boldsymbol{SM}(\boldsymbol{T_1}, \boldsymbol{P_1}) = (\boldsymbol{1} - \boldsymbol{\alpha}) * \boldsymbol{S_o}(\boldsymbol{T_1}, \boldsymbol{P_1}) + \boldsymbol{\alpha} * \boldsymbol{S_r}(\boldsymbol{T_1}, \boldsymbol{P_1}), \quad (15)$$

TABLE 1 Test accuracy of COVID-Net with COVID-chestxray

Backbone (%)	Adam	SGD	AdaBound	Apollo	ACQN-H
VGG16	95.44	95.44	96.25	96.36	96.74
ResNet50	94.41	94.19	95.49	95.44	95.60
DenseNet121	95.17	95.27	95.82	96.36	96.52

where $\alpha \in [0, 1]$, S_o is the object-aware similarity and S_r represents the region-aware similarity.[38]

3. Enhanced-alignment measure:

$$EM(T_1, P_1) = \frac{|\phi(T_1, P_1)|}{w \times h}, \tag{16}$$

where w and h are the width and height of the input CT image, respectively, and $\phi(\cdot)$ represents the enhanced alignment matrix.[39]

4. Mean absolute error:

$$MAE(T_1, P_1) = \frac{||T_1 - P_1||_1}{w \times h}. \tag{17}$$

5. Intersection over union:

$$IoU(T_1, P_1) = \frac{|T_1 \cap P_1|}{|T_1 \cup P_1|}. \tag{18}$$

6. True positive rate:

$$TPR(T_1, P_1) = \frac{|T_1 \cap P_1|}{|T_1 \cap P_1| + |T_0 \cap P_0|}. \tag{19}$$

7. True negative rate:

$$TNR(T_0, P_0) = \frac{|T_0 \cap P_0|}{|T_0|}. \tag{20}$$

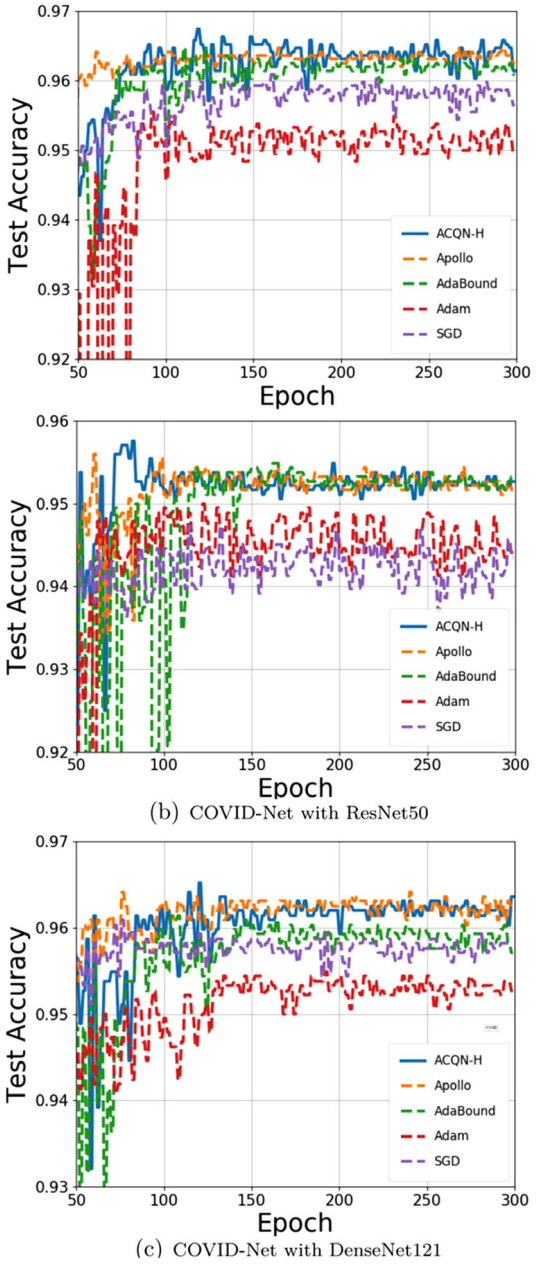

FIGURE 1 Test accuracy curves of COVID-Net on COVID-chestxray using VGG16, ResNet50, and DenseNet121 as backbone

3 | RESULTS

3.1 | Detection of COVID-19

To assess the generalization performance of ACQN-H on medical image classification, we also use COVID-Net with VGG16, ResNet50, and DenseNet121 as backbones on the COVID-chestxray dataset, and results are shown in Table 1, ACQN-H outperforms other optimizers on classification accuracy in all experiments and achieves 0.38%, 0.16%, and 0.16% higher accuracy than Apollo with the backbones VGG16, ResNet50, and DenseNet121, respectively. Test accuracy curves are reported in Figure 1. As can be seen, the test accuracy of ACQN-H is better than that of other optimizers. Moreover, fivefold cross-validation results using ACQN-H are reported in Table 2. It can be found that the average Dice values among different folds vary slightly.

3.2 | COVID-19 lung infection segmentation

In the COVID-19 lung infection segmentation task, we experiment on the Inf-Net model with the COVID-CT dataset. The performance of ACQN-H is evaluated through six widely adopted metrics, that is, Dice, TPR,

TABLE 2 Quantitative results of fivefold cross-validation using COVID-Net with COVID-chestxray

	Validation set			Test Set		
Backbone(%)	**VGG16**	**ResNet50**	**DenseNet121**	**VGG16**	**ResNet50**	**DenseNet121**
Fold-0	99.01	98.95	99.21	97.21	95.95	97.61
Fold-1	98.55	98.15	98.65	96.05	95.15	95.97
Fold-2	97.62	98.02	98.82	95.92	95.32	96.32
Fold-3	97.86	97.36	97.93	97.86	95.36	97.13
Fold-4	98.66	97.72	98.54	96.66	96.22	95.57
Avg	98.34	98.04	98.63	96.74	95.60	96.52

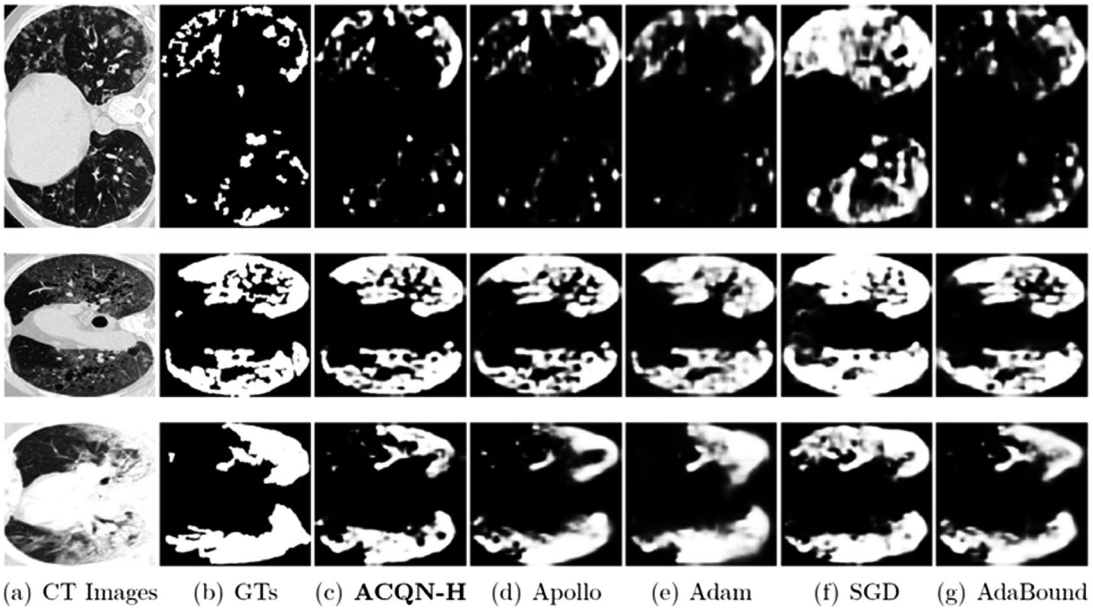

FIGURE 2 Segmentation results on COVID-CT with Inf-Net

TABLE 3 Assessment of Inf-Net with COVID-CT. (For metric MAE, lower is better, for other metrics, higher is better)

(%)	Adam	SGD	AdaBound	Apollo	ACQN-H
Dice	68.7	57.4	64.8	69.0	70.0
TPR	68.1	79.2	65.9	68.2	68.4
TNR	94.9	85.7	94.3	95.6	95.4
SM	76.5	63.8	75.0	76.9	76.0
EM	84.9	72.4	83.8	87.3	88.7
MAE	7.7	14.9	8.9	7.7	7.4

TABLE 4 Quantitative results of fivefold cross-validation using COVID-CT with Inf-Net

%	Validation set	Test set
Fold-0	78.50 ± 0.10	68.51 ± 0.57
Fold-1	77.90 ± 0.14	70.87 ± 0.47
Fold-2	79.71 ± 0.10	65.45 ± 0.90
Fold-3	80.17 ± 0.10	73.31 ± 0.65
Fold-4	81.77 ± 0.07	71.86 ± 0.62
Avg.	79.61 ± 0.11	70.02 ± 0.59

TNR, SM, EM, and MAE. For metric MAE, lower is better and for other metrics, higher is better.

Table 3 presents quantitative results of COVID-19 lung infection segmentation. It shows that ACQN-H has the best scores in terms of Dice, TPR, EM, and MAE. Particularly, ACQN-H achieves 1.0% better Dice as compared to Apollo. Figure 2 also gives some visual comparison examples of COVID-19 lung infection segmentation. Comparing with other optimizers, ACQN-H yields infection segmentation results with more accurate boundaries. Besides, Table 4 shows quantitative fivefold validation results of COVID-19 lung infection segmentation in terms of Dice. As can be seen, the variance on the test set is much larger than on the validation set. Moreover, the performance on the test set drops about 9% on average. The potential reason could be the small size of the dataset.

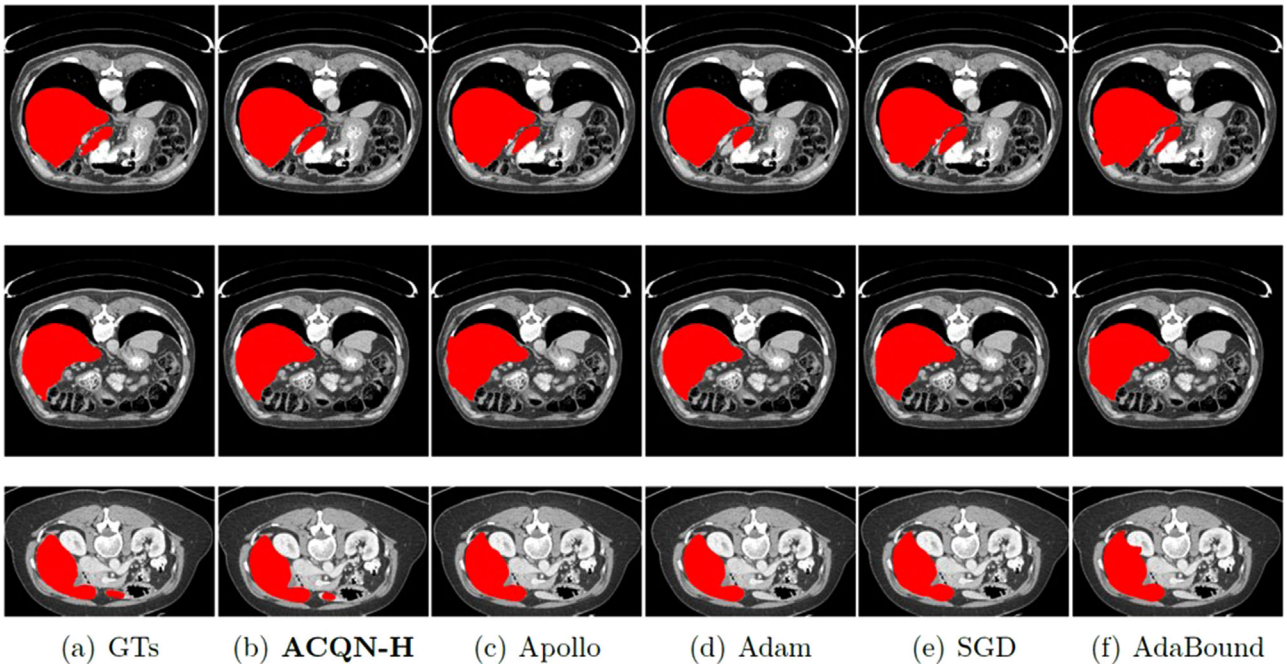

FIGURE 3 Segmentation results on LiTS2017 with ResUNet. The red pixels denote the liver region.

TABLE 5 Assessment of ResUNet on LiTS2017. (For Dice, higher is better, while for other evaluation metrics, lower is better)

(%)	Adam	SGD	AdaBound	Apollo	ACQN-H
Dice	91.22	90.87	90.62	91.20	**93.46**
TPR	97.53	98.48	98.40	**98**.77	**98**.77
TNR	86.90	84.77	84.80	85.37	**89.26**
IoU	84.43	83.58	83.20	84.14	**88.03**

3.3 | Liver tumor segmentation

We experiment with the ResUNet model on the LiTS2017 dataset and evaluate the results with assessments Dice, IoU, TPR, and TNR. For all these evaluation metrics, higher is better.

Quantitative results of liver tumor segmentation are shown in Table 5. As can be seen, ACQN-H achieves the best results and outperforms Apollo in terms of Dice by 2.3%. Figure 3 gives some visual comparison examples of liver tumor segmentation. We can see that the segmentation results using ACQN-H are more similar to the GTs than other optimizers. Besides, quantitative fivefold validation results of liver tumor segmentation in terms of Dice are shown in Table 6. As can be seen, the variance among results in each fold is low. Moreover, the relative performance on the test set drops only about 2.14%.

3.4 | Optic disc/cup segmentation

In the optic disc/cup segmentation task, we experiment on the RIGA dataset with the MRNet model, and the performance is evaluated through IoU and Dice. We present quantitative experiments in Table 7. Obviously, ACQN-H has advantage on cup segmentation and achieves 0.6% and 1.3% better scores for Dice and IoU, respectively, when comparing with Apollo. ACQN-H also achieves comparable results on disc segmentation. Figure 4 presents some visualized segmentation results. From the first row, all optimizers achieve excellent performance on optic disc segmentation, and their results are similar to the GT. Meanwhile, the second row shows that the optic cup segmentation result using ACQN-H achieves superior performance.

TABLE 6 Quantitative results of fivefold Cross-validation using ResUNet with LiTS2017

%	Validation set	Test set
Fold-0	95.03 ± 0.09	94.04 ± 0.04
Fold-1	95.12 ± 0.09	93.13 ± 0.04
Fold-2	96.68 ± 0.09	94.01 ± 0.04
Fold-3	95.45 ± 0.09	93.11 ± 0.04
Fold-4	95.72 ± 0.09	93.01 ± 0.04
Avg	95.60 ± 0.09	93.46 ± 0.04

TABLE 7 Assessment of MRNet with RIGA. (Higher is better)

(%)	Adam	SGD	AdaBound	Apollo	ACQN-H
Disc Dice	94.9	**97**.7	97.5	97.6	97.6
Disc IoU	90.5	**95.5**	95.1	95.3	95.4
Cup Dice	83.3	83.3	81.9	83.2	**83.8**
Cup IoU	73.2	73.2	71.5	72.9	**74.2**

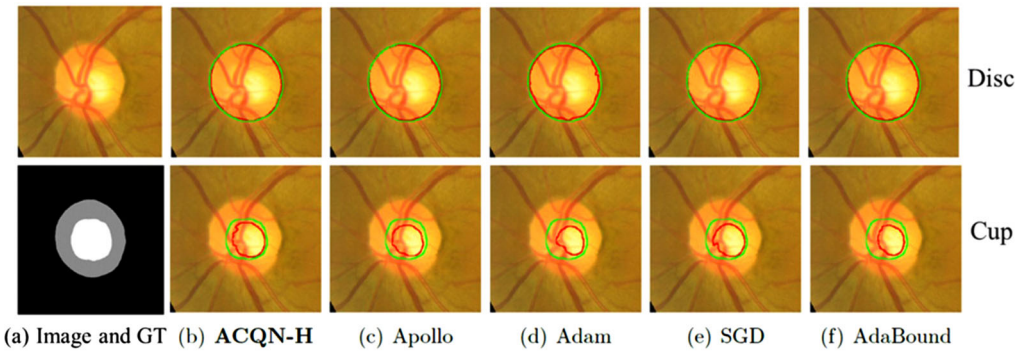

FIGURE 4 A sample of segmentation result on RIGA with MRNet. (a) represents the original image (above) and ground truth (below). The segmentation boundaries of GT (green) and the predicted optic disc (red) for different optimizers are shown in the first row of (b)–(f), while the results of cup segmentation are shown in the second row.

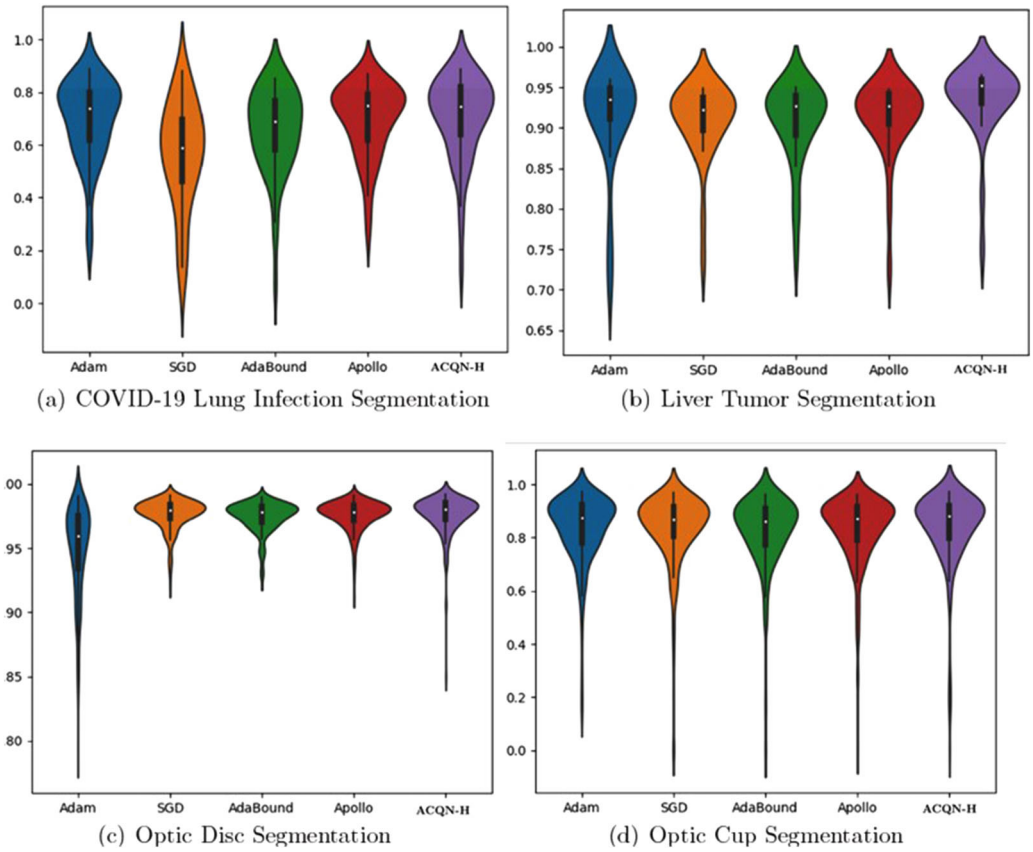

FIGURE 5 The violin plots present the dice of different optimizers for COVID-19 lung infection segmentation, liver tumor segmentation, and optic disc/cup segmentation.

We also present quantitative fivefold validation results of optic disc/cup segmentation in Table 8. It can be found that optic disc segmentation can achieve high performance with low variance. However, the performance of optic cup segmentation is not as good as optic disc segmentation, and the average Dice drops about 14% on the test set with obvious larger variance. This indicates that optic cup segmentation remains a challenging problem.

4 | DISCUSSION

In this paper, we present a novel ACQN-H for medical image analysis. Our method can capture the curvature of the loss function by diagonally approximated Hessian and the norm of difference between previous two estimates. Additionally, ACQN-H hires high-order moment through exponential moving average on iteratively calculated Hessian approximations. The method can help

TABLE 8 Quantitative results of fivefold Cross-validation using MRNet with RIGA

Subtask (%)	Optic disc		Optic cup	
	Validation set	Test set	Validation set	Test Set
Fold-0	99.01 ± 0.02	97.81 ± 0.02	88.36 ± 0.14	84.24 ± 0.68
Fold-1	98.70 ± 0.03	97.15 ± 0.02	88.18 ± 0.14	82.31 ± 0.71
Fold-2	98.57 ± 0.01	96.47 ± 0.02	86.97 ± 0.10	83.55 ± 0.62
Fold-3	99.25 ± 0.01	98.32 ± 0.02	86.13 ± 0.14	83.97 ± 0.53
Fold-4	99.07 ± 0.01	98.25 ± 0.01	87.61 ± 0.17	84.93 ± 0.57
Avg	98.92 ± 0.02	97.60 ± 0.02	87.45 ± 0.15	83.81 ± 0.64

to escape from saddle points more efficiently and train a DNN with better performance.

ACQN-H is evaluated through a wide range of medical image analysis tasks using state-of-the-art models. These include COVID-chestxray for the detection of COVID-19 using COVID-Net, COVID-CT for COVID-19 lung infection segmentation using Inf-Net, LiTS2017 for liver tumor segmentation using ResUNet and RIGA for optic disc/cup segmentation using MRNet. For the detection of COVID-19, quantitative results are reported in Table 1. As a medical image classification task, ACQN-H consistently achieves the best accuracy with different backbones. Test accuracy curves are reported in Figure 1, from which we can see that ACQN-H has a better convergence than Adam, SGD, and AdaBound. Quantitative results of COVID-19 lung infection segmentation are shown in Table 3, from which ACQN-H achieves the best score in terms of dice, EM, and MAE, only slightly inferior to Apollo in SM and TNR. This implies the segmentation results from ACQN-H are more similar to GTs when comparing with Adam, SGD, and AdaBound. In terms of TPR and TNR, the resulting image of ACQN-H has the least proportion of missegmented pixels. In addition, Figure 2 also gives some visual comparison examples. The result of ACQN-H segment the regions infected with COVID-19 more accurately. Results of liver tumor segmentation are shown in Table 5 and Figure 3. Obviously, ACQN-H outperforms other optimizers in terms of quantization and segmentation. As quantitative results of optic disc/cup segmentation shown in Table 7, ACQN-H achieves the best dice and IoU on optic cup segmentation, which implies that ACQN-H can better segment optic cups as GT. For optic disc segmentation, ACQN-H is only slightly lower than Adam which is better than other optimizers. Figure 4 gives some visualized segmentation results. ACQN-H has a visually better optic cup segmentation result, while its disc segmentation result is comparable.

We find that the segmentation results do not match the GTs well on COVID-19 lung infection segmentation and optic cup segmentation. This has resulted from two main reasons. At first, the datasets COVID-CT and RIGA have limited labeled cases. Second, the test samples are complex from a visual point. For example, the samples in the COVID-19 lung infection segmentation task include many small infections with irregular margins. For the optic cup segmentation task, the cases fail most often due to a weak boundary between the optic disc and optic cup. Thus, these two tasks are much more challenging, while ACQN-H still achieves the best results in visual.

Figure 5 shows the violin plots of COVID-19 lung infection, liver tumor, and optic disc/cup segmentation results using different optimizers in terms of Dice. It gives the summary statistics and the entire distribution of the quantitative results. As can be seen, ACQN-H achieves the best lower quartile, median, and upper quartile in all tasks, which indicates that most cases segmented using ACQN-H get higher Dice.

5 | CONCLUSION

We have proposed ACQN-H, a novel and efficient adaptive cubic quasi-Newton optimizer with a high-order moment for medical image analysis, and its superiority is demonstrated on four types of datasets. ACQN-H only requires at most first-order gradients and updates with linear complexity for both time and memory, thus it is quite suitable for large-scale deep learning based medical image analysis and is expected to boost the performance of existing DNNs for medical image analysis.

ACKNOWLEDGMENTS

This research was supported in part by the Hunan Provincial Natural Science Foundation (Grants No. 2019JJ50746) and the National Natural Science Foundation of China (Grants No. 61602494).

CONFLICT OF INTEREST

The authors have no conflict to disclose.

REFERENCES

1. Park S, Do K, Kim S. What should medical students know about artificial intelligence in medicine? *J Educ Eval Health Prof*. 2019;16:18.
2. Cai L, Gao J, Zhao D. A review of the application of deep learning in medical image classification and segmentation. *Ann Transl Med*. 2020;8(11):713.
3. Sze V, Chen Y, Yang T, Emer JS. Efficient processing of deep neural networks: A tutorial and survey. *Proc IEEE*. 2017;105(12):2295-2329.
4. Ronneberger O, Fischer O, Brox T. *U-net: Convolutional Networks for Biomedical Image Segmentation*. Springer International Publishing; 2015.
5. Li X, Chen H, Qi X, et al. H-denseunet: Hybrid densely connected Unet for liver and liver tumor segmentation from CT volumes. *IEEE Trans Med Imaging*. 2017;37(12):2663-2674.
6. Xu W, Liu H, Wang X, et al. Liver segmentation in CT based on ResUNet with 3D probabilistic and geometric post process. In: *IEEE International Conference on Signal Image Processing*. IEEE; 2019:685-689.

7. Fausto M, Navab N, Ahmadi SA. V-net: fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE; 2016:565-571.
8. Zhou Z, Siddiquee M, Tajbakhsh N, et al. Unet++: A nested u-net architecture for medical image segmentation. In: *4th Deep Learning in Medical Image Analysis (DLMIA) Workshop*. Springer; 2018:3-11.
9. Wei J, Shuang Y, Junde W, et al. Learning calibrated medical image segmentation via multi-rater agreement modeling. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE; 2021:12341–12351.
10. Islam MM, Karray F, Alhajj R, et al. A review on deep learning techniques for the diagnosis of novel coronavirus (covid-19). *IEEE Access*. 2021;9:30551-30572.
11. Wang L, Lin ZQ, Wong A. Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest X-ray images. *Sci Rep*. 2020;10:19549.
12. Gunraj H, Wang L, Wong A. COVIDNet-CT: A tailored deep convolutional neural network design for detection of covid-19 cases from chest ct images. *Front Med*. 2020;7:608525.
13. Kim M, Kang J, Kim D, etc. Hi-covidnet: Deep learning approach to predict inbound COVID-19 patients and case study in South Korea. In: *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Virtual Event, CA, August 23-27. ACM;2020:3466-3473.
14. Gunraj H, Wang L, Wong A. Mini-COVIDNet: Efficient light weight deep neural network for ultrasound based point-of-care detection of covid-19. *Front Med*. 2020;68:2023-2037.
15. Fan DP, Zhou T, Ji GP, et al. Inf-Net: automatic covid-19 lung infection segmentation from CT images. *IEEE Trans Med Imaging*. 2020:39(8):2626-2637.
16. Zhuang J, Tang T, Tatikonda S, et al. Adabelief optimizer: adapting stepsizes by the belief in observed gradients. In: *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*. Curran Associates; 2020:18795-18806.
17. Ruder S. An overview of gradient descent optimization algorithms. In: *CoRR*. *arXiv 1609.04747*. 2016. http://arxiv.org/abs/1609.04747
18. Robbins H, Monro S. A stochastic approximation method. *Ann Math Stat*. 1951:22(3):400-407.
19. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, May 7-9, 2015, and Conference Track Proceedings*. 2015;abs/1412.6980.
20. Loshchilov I, Hutter F. Decoupled weight decay regularization. In *7th International Conference on Learning Representations, ICLR 2019*, OpenReview.net, New Orleans, LA, May 6-9, 2019. 2019.
21. Luo L, Xiong Y, Liu Y, Sun X. Adaptive gradient methods with dynamic bound of learning rate. In: *7th International Conference on Learning Representations, ICLR 2019*, OpenReview.net, New Orleans, LA. 2019.
22. Liu L, Jiang H, He P, et al. On the variance of the adaptive learning rate and beyond. *CoRR*. 2019;abs/1908.03265.
23. Li W, Zhang Z, Wang X, Luo P. Adax: adaptive gradient descent with exponential long term memory. *CoRR*. 2020;abs/2004.09740.
24. Yao Z, Gholami A, Shen S, Keutzer K, Mahoney MW. Adahessian: an adaptive second order optimizer for machine learning. *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2–9, 2021*. AAAI Press, 2021:10665-10673.
25. Avron H, Toledo S. Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix. *J ACM*. 2011;58:8.
26. Keskar NS, Berahas AS. AdaQN: an adaptive quasi-Newton algorithm for training RNNs. In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*, December 4-9, 2017, Curren Associates; 2016:4148-4158.
27. Wang X, Ma S, Goldfarb D, Liu W. Stochastic quasi-Newton methods for nonconvex stochastic optimization. *SIAM J Optim*. 2017;27(2):927-956.
28. Ma X. Apollo: an adaptive parameter-wise diagonal quasi-Newton method for nonconvex stochastic optimization. *CoRR*. 2020;abs/2009.13586.
29. Nesterov YE, Polyak BT. Cubic regularization of Newton method and its global performance. *Math Program*. 2006;108:177-205.
30. Zhu M, Nazareth JL, Wolkowicz H. The quasi-Cauchy relation and diagonal updating. *SIAM J Optim*. 1999;9:1192-1204.
31. Cohen JP, Morrison P, Dao L, Roth K, Duong TQ, Ghassemi M. Covid-19 image data collection: prospective predictions are the future. *arXiv 2006.11988*. 2020.
32. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: *3rd International Conference on Learning Representations, ICLR 2015*, OpenReview.net, San Diego, CA, USA: Conference Track Proceedings; 2015.
33. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In; *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*. IEEE; 2016:770-778.
34. Huang G, Liu Z, Weinberger KQ. Densely connected convolutional networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Honolulu, HI. IEEE; 2017:2261-2269.
35. Yang X, He X, Zhao J, Zhang Y, Zhang S, Xie P. Covid-CT-dataset: A CT scan dataset about covid-19. *arXiv 2003.13865*. 2020. https://arxiv.org/abs/2003.13865
36. Bilic P, Christ P, Vorontsov E, Chlebus G, Wu J. The liver tumor segmentation benchmark (lits). In: *CVPR2019. arXiv 1901.04056*. 2019. http://arxiv.org/abs/1901.04056
37. Almazroa A, Alodhayb S, Osman E, et al. Agreement among ophthalmologists in marking the optic disc and optic cup in fundus images. *Int Ophthalmol*. 2017;37(3):701-717.
38. Fan DP, Cheng MM, Liu YTL, Borji A. Structure-measure: a new way to evaluate foreground maps. In: *Structure-Measure: A New Way to Evaluate Foreground Maps*. IEEE Computer Society; 2017:4548-4557.
39. Fan DP, GONG C, Cao Y, et al. Enhanced-alignment measure for binary foreground map evaluation. In: *The 27th International Joint Conference on Artificial Intelligence*. ijcai.org; 2018:698-704.

How to cite this article: Liu Y, Zhang M, Zhong Z, Zeng X. A novel adaptive cubic quasi-Newton optimizer for deep learning based medical image analysis tasks, validated on detection of COVID-19 and segmentation for COVID-19 lung infection, liver tumor, and optic disc/cup. *Med Phys*. 2022;1-11.
https://doi.org/10.1002/mp.15969