



Published in final edited form as:

*Med Image Comput Assist Interv.* 2019 October ; 11769: 284–292.

doi:10.1007/978-3-030-32226-7\_32.

## Targeting Precision with Data Augmented Samples in Deep Learning

**Pietro Nardelli, Raúl San José Estépar**

Applied Chest Imaging Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

### Abstract

In the last five years, deep learning (DL) has become the state-of-the-art tool for solving various tasks in medical image analysis. Among the different methods that have been proposed to improve the performance of Convolutional Neural Networks (CNNs), one typical approach is the augmentation of the training data set through various transformations of the input image. Data augmentation is typically used in cases where a small amount of data is available, such as the majority of medical imaging problems, to present a more substantial amount of data to the network and improve the overall accuracy. However, the ability of the network to improve the accuracy of the results when a slightly modified version of the same input is presented is often overestimated. This overestimation is the result of the strong correlation between data samples when they are considered independently in the training phase. In this paper, we emphasize the importance of optimizing for accuracy as well as precision among multiple replicates of the same training data in the context of data augmentation. To this end, we propose a new approach that leverages the augmented data to help the network focus on the precision through a specifically-designed loss function, with the ultimate goal to improve both the overall performance and the network's precision at the same time. We present two different applications of DL (regression and segmentation) to demonstrate the strength of the proposed strategy. We think that this work will pave the way to a explicit use of data augmentation within the loss function that helps the network to be invariant to small variations of the same input samples, a characteristic that is always required to every application in the medical imaging field.

### Keywords

Deep learning; Data augmentation; Accuracy; Precision

## 1 Introduction

Thanks to their demonstrated great performance and potential, deep learning (DL) techniques are quickly becoming the main tools for medical imaging analysis. In recent years, several new architectures and DL applications have been proposed in the literature trying to solve various problems [5,9]. However, depending on the complexity of the task

that a DL model has to perform, the number of parameters that need to be tuned to reduce the loss might be very high.

Therefore, in order to get good performance, a proportionally high amount of examples should be shown to the network, as a model's performance typically increases as the quality, diversity, and the amount of training data increases. However, especially in medical imaging tasks, collecting enough data is often highly complicated and one of the biggest problems developers have to face when creating a new DL tool is over-fitting to the training dataset, which happens when the network is robust with the shown examples, but it is not able to generalize well the output if initial conditions are even slightly modified.

The most typical approach used to overcome this issue is the augmentation of the training data-set by means of various synthetic transformations of the input image, a method that is particularly helpful when only a small amount of data is available, and that may help increase the level of invariance of the network to known transformation of the input image space.

In recent years, much effort has been spent trying to improve and automatize augmentation techniques in order to get the network to learn varying conditions and improve the overall accuracy and robustness [2]. Recently, adversarial neural networks (GANs) have been proposed [4], and one possible application domain for these techniques is data augmentation [1, 3].

However, in addition to accuracy and robustness an important metric that should also be taken into account, especially when trying to solve medical imaging issues, is precision. Precision is a metric that indicates that the network is able to output the same values when replicas of the same input, although slightly modified, are presented. In the majority of proposed works though, this aspect is underestimated, and the overall performance is the only metric considered during the training phase of the network.

In this paper, we present a novel approach to augmented samples that takes into account both accuracy and precision during the training phase. We propose to use a specifically-designed loss function that can be modified according to the problem into consideration and combines both metrics to improve the network ability to learn image characteristics regardless of varying conditions properly.

To demonstrate the strength of the proposed strategy, two applications of DL with and without the presented approach are shown. First, a regression problem where 25 modified replicas of the same input are used to measure airway lumen radius on CT images accurately is presented. Then, a segmentation of the lung region is shown to demonstrate the improvement achieved both in accuracy and precision when 20 synthetically modified versions of each input example are used.

## 2 Material and Methods

In order to properly target precision for DL techniques, we propose the development of a customized loss function that can be designed around the problem of interest to take into

account augmented data and simultaneously improve the overall accuracy as well as the network's precision. The basic idea is that when a synthetically modified version of the same input is presented to the network, the exact same output should be expected. In this section, we first introduce the general loss, and we then present two different versions of it as we adapt it to two separate problems.

## 2.1 Accuracy-Precision Loss

In this paper, we propose the usage of a new loss for DL algorithms that consists of two terms; a term that represents the loss of the accuracy over all images,  $\mathcal{L}_\mu$ , and a second term that takes into account the precision over the synthetically augmented replicas of the original inputs,  $\mathcal{L}_\sigma$ .

If we represent all input and augmented images in a  $N \times M$  matrix where the first column contains the  $N$  original inputs and all other columns contain  $M$  augmented replicas (one column for each augmentation) then, the precision loss,  $\mathcal{L}_\sigma$ , is computed over each row  $N$  of the matrix, while the accuracy loss,  $\mathcal{L}_\mu$ , is given over all  $N \times M$  inputs, as shown in Fig. 1. More formally, the accuracy-precision loss,  $\mathcal{L}_{AP}$ , is given by the following expression:

$$\mathcal{L}_{AP}(\mathbf{y}, \hat{\mathbf{y}}) = \omega \cdot \mathcal{L}_\mu(\mathbf{y}, \hat{\mathbf{y}}) + \lambda \cdot \mathcal{L}_\sigma(\mathbf{y}, \hat{\mathbf{y}}) \quad (1)$$

where  $\mathbf{y}$  is the true value of the input,  $\hat{\mathbf{y}}$  is the predicted value, and  $\omega, \lambda$  define how the two terms of the loss are weighted. From this it follows that:

$$\begin{aligned} \mathcal{L}_{AP}(\mathbf{y}, \hat{\mathbf{y}}) = & \omega \cdot \frac{1}{N \times M} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \mathcal{L}_\mu(y_{i,j}, \hat{y}_{i,j}) + \lambda \\ & \cdot \frac{1}{N} \sum_{i=0}^{N-1} \left( \frac{1}{M} \sum_{j=0}^{M-1} \mathcal{L}_\sigma(y_{i,j}, \hat{y}_{i,j}) \right) \end{aligned} \quad (2)$$

Depending on the specific task that the deep learning model has to perform,  $\mathcal{L}_\mu$  and  $\mathcal{L}_\sigma$  can be specifically designed according to Eq. 2. In the following two sections we show two examples of accuracy-precision loss as it might adapted to regression and segmentation tasks.

## 2.2 Accuracy-Precision Loss for Regression

In this section, we present an accuracy-precision loss applied to a regression task to measure the lumen nominal radius of small airways from chest CT images. For this problem, we developed a model-generator that creates 2D synthetic patches of  $32 \times 32$  pixels (with a resolution of 0.5mm) on the reformatted axial plane that resemble patches around real airways.

Thanks to this generative model, training examples are not an issue, as they can be generated at will with known lumen size. The regressor consists of a 9-layer 2D network with seven convolutional and two fully-connected layers, as proposed in [6], and is trained on 2,500,000 synthetic patches, generated as 100,000 individual images for which 25 augmented replicas

are created by maintaining the same lumen and wall thickness and varying point spread function (PSF), rotation, flipping, and translation, and adding a various number of additional vessels around the airway.

While trying to reduce the relative error (RE) over all images (accuracy), we want the regressor to also learn that the same measurement for airway lumen radius should be provided across all replicas regardless of possible confounding factors inside the patch. Therefore, the accuracy loss is given by the RE:

$$\mathcal{L}_\mu(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{N \times M} \cdot \sum_{i=0, j=0}^{N, M} \frac{|y_{i,j} - \hat{y}_{i,j}|}{y_{i,j}} \quad (3)$$

where  $\mathbf{y}$  indicates the ground-truth measure of a synthetic patch provided by the generative model,  $\hat{\mathbf{y}}$  the measure predicted by the regressor,  $N$  is the total number of individual patches ( $N=100,000$ ), and  $M$  is the number of replicas ( $M=25$ ). Conversely, for the precision loss we want to minimize the error across the  $M$  replicas by:

$$\mathcal{L}_\sigma(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{N} \cdot \sum_{i=0}^N \left( \frac{1}{M} \cdot \sum_{j=0}^M (y_{i,j} - \hat{y}_{i,j})^2 - \left( \frac{1}{M} \cdot \sum_{j=0}^M (y_{i,j} - \hat{y}_{i,j}) \right)^2 \right) \quad (4)$$

Moreover, in this specific case we want to give more importance to precision with respect to accuracy. Therefore, we set  $\omega = 1$  and  $\lambda = 2$  and the final accuracy-precision loss for the presented regression problem is given by:

$$\begin{aligned} \mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = & \frac{1}{N \times M} \cdot \sum_{i=0, j=0}^{N, M} \frac{|y_{i,j} - \hat{y}_{i,j}|}{y_{i,j}} + 2.0 \cdot \frac{1}{N} \\ & \cdot \sum_{i=0}^N \left( \frac{1}{M} \cdot \sum_{j=0}^M (y_{i,j} - \hat{y}_{i,j})^2 - \left( \frac{1}{M} \cdot \sum_{j=0}^M (y_{i,j} - \hat{y}_{i,j}) \right)^2 \right) \end{aligned} \quad (5)$$

### 2.3 Accuracy-Precision Loss for Segmentation

The same concept can also be applied to a segmentation task. In this paper, we use a 2D intra-parenchymal lung segmentation from CT images to show a different application domain for the proposed technique. Since our final goal is to show the ability of our method to improve results compared to traditional approaches, and not to obtain an optimal segmentation, we used a modified version of U-Net [7] as segmenter.

In this case, as ground-truth we used axial lung labels obtained with the method described in [8] and visually evaluated for correctness. We randomly extracted 2,100 2D axial CT slices and corresponding label maps from 300 cases from high-resolution chest CT scans of the COPDGene study phase 2.

For memory reasons, we resampled images to a size of  $256 \times 256$  pixels. To create the augmented data, we generated 10 replicas for each original image by introducing random noise, rotation, flipping, translation, and skew, and shear transformations. As a main metric

for the segmentation, we used the dice coefficient score to compare the label produced by our segmenter to the ground-truth. However, while we want the dice coefficient as high as possible over all images, we also want an as small as possible standard deviation of the error across the 10 image replicas. To this end, we used an accuracy loss given by:

$$\mathcal{L}_\mu(\mathbf{y}, \hat{\mathbf{y}}) = 1.0 - 2.0 \cdot \frac{|\mathbf{y} \cap \hat{\mathbf{y}}|}{|\mathbf{y}| + |\hat{\mathbf{y}}|} \quad (6)$$

where  $\mathbf{y}$  indicates the ground-truth label,  $\hat{\mathbf{y}}$  the label provided by the network,  $N=2,100$  is the total number of individual images, and  $M=10$  is the number of synthetically augmented replicas. In this case, the precision loss is computed as the standard deviation of the dice coefficient loss over the  $M=10$  replicas:

$$\mathcal{L}_\sigma(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{N} \cdot \sum_{i=0}^N \sqrt{\text{VAR}_M \left( 1.0 - 2.0 \cdot \frac{|\mathbf{y} \cap \hat{\mathbf{y}}|}{|\mathbf{y}| + |\hat{\mathbf{y}}|} \right)} \quad (7)$$

where  $\text{VAR}_M$  indicates the variance across the  $M$  replicas. As in the regression case, we set  $\omega = 1$  and  $\lambda = 3.0$  to obtain the global accuracy-precision loss as:

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = 1.0 - 2.0 \cdot \frac{|\mathbf{y} \cap \hat{\mathbf{y}}|}{|\mathbf{y}| + |\hat{\mathbf{y}}|} + 3.0 \cdot \frac{1}{N} \cdot \sum_{i=0}^N \sqrt{\text{VAR}_M \left( 1.0 - 2.0 \cdot \frac{|\mathbf{y} \cap \hat{\mathbf{y}}|}{|\mathbf{y}| + |\hat{\mathbf{y}}|} \right)} \quad (8)$$

### 3 Experimental Setup

To evaluate the proposed approach, we compared the results obtained for the regression and the segmentation tasks with the presented accuracy-precision loss and its accuracy-only version.

For airway lumen measurement, we first used 200,000 randomly generated synthetic patches to evaluate the overall mean RE. Then, we created a dataset by varying the level of noise in the range  $\sigma_n \in [0, 40]$  HU (smoothing level fixed at 1.3mm) and generating 100 synthetic replicas for each noise value, and a second dataset fixing the level of noise at 25 HU and changing the applied smoothness in the range  $\sigma_s \in [0.4, 0.9]$ mm to generate 100 synthetic replicas per degree of smoothing.

To create the two datasets, we fixed the wall thickness at 1.5mm and used three airway lumen values (small: 0.5mm; medium: 2.5mm; large: 4.5mm) randomly varying all other parameters of the geometric model. For these two datasets, we computed the mean RE (in %) across the 100 patches for each level of noise and smoothness, to demonstrate that the proposed loss function helps improve not only the accuracy but also the precision of network when initial conditions change. For the accuracy-only method, we used a simple mean RE loss.

On the other hand, to demonstrate the improved ability to segment the lung region of a U-net that uses the proposed loss function, we used the 55 cases from the Lobe and Lung Analysis 2011 (LOLA11) challenge that were first segmented with the method presented in [8], and

were then visually inspected and manually refined. We used the Dice coefficient score for the comparison. The same process was finally repeated using the same neural network architecture, but only with a Dice coefficient loss.

## 4 Results

The mean RE when using the proposed accuracy-precision loss to measure the airway lumen was 2.06%, while when a mean RE loss is used the mean RE is 3.01%. Results obtained when fixing three values of airway lumen (0.5, 2.5, and 4.5mm) and varying the level of noise and smoothing for the two methods are presented in Fig. 2. As shown, the RE obtained with both methods and for both measurements is stable across the different levels of noise and smoothness.

However, while with the proposed loss a very high accuracy is obtained (RE close to 0 for large and medium structures, RE around -10 for small airways), using a traditional loss function a bigger RE is obtained for all structures. Also, the standard deviation is much smaller when a precision-accuracy loss function is used. This effect is visible when varying both the noise (first row in Fig. 2) and the smoothness (second row in Fig. 2).

For the segmentation analysis, the mean Dice score -in comparison with the method proposed in [8]- when using the loss function presented in Eq. 8 was  $97.6\% \pm 0.027$ , while a traditional loss is used, a mean Dice score of  $95.8\% \pm 0.029$  was obtained.

Dice coefficient results stratified by lung volume (low and high volume defined as being below or above the median lung volume of the LOLA datasets) and by lung size (right vs. left) are presented in Fig. 4. As shown, while the Dice scores are consistent for both lungs and by level of volume lung for both methods independently, the proposed loss has a consistent higher Dice score ( $p < 0.001$ ) with a lower variance implying a more robust result.

From an accurate analysis of segmentation of single cases, it was clear that with a traditional loss the segmentation has a higher tendency to leak inside the trachea, whereas the proposed loss function seems to help the network segment the lung region in a more accurate and precise way (see Fig. 3).

## 5 Discussion and Conclusion

In this paper, we presented a novel approach to data augmentation for deep learning tasks that instead of focusing only on the overall performance of the network may help increase both accuracy and precision of a neural network.

The modified data is used not only to increase the amount of data presented to the network but also to reduce the variance of the output when similar inputs are presented. To this end, we introduce a new generalized loss function that leverages the information provided by the augmented data and can be adapted to the specific problem of interest. Results from the application of the loss function to regression and segmentation tasks showed that while the presented technique helps improve the overall accuracy, the precision of the DL model is also increased.

For the lung segmentation task, we want to point out that the final goal was not to obtain an optimal segmentation, but to demonstrate that if the proposed function is utilized, results improve compared to the usage of a traditional loss. For future work, we will validate the proposed loss function by showing that the best performing deep neural network reported for LOLA11 trained with the proposed loss function improves the final segmentation. Also, a classification task will also be considered for further testing.

The main limitation of the proposed loss function is given by the fact that when big image data is required the batch size will necessarily get smaller, significantly reducing the number of replicas that can be shown to the network.

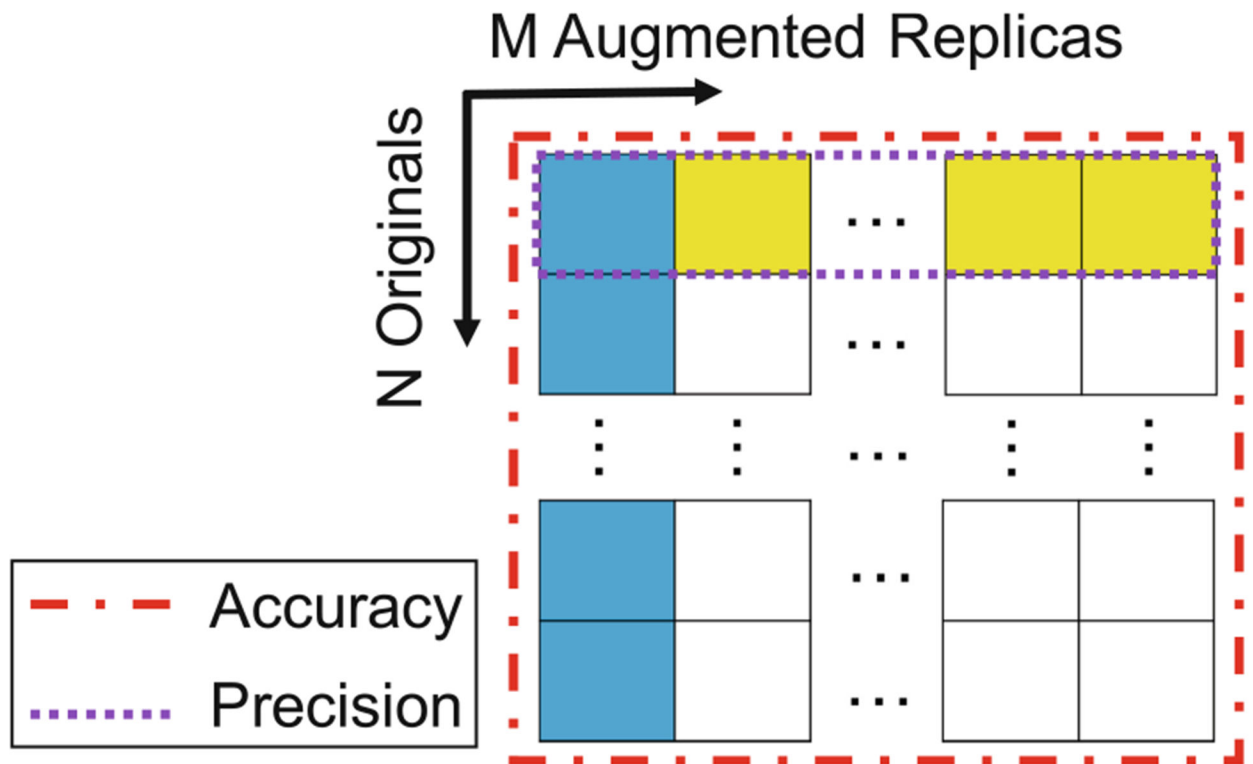
In conclusion, we believe that the proposed approach to data augmentation will pave the way to a new way of using the synthetically modified data and will help to improve the performance of DL techniques in a field where a significant amount of data is not available at no cost.

## Acknowledgments

This work has been partially funded by the National Institutes of Health NHLBI awards R01HL116931, R01HL116473, and R21HL14042. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the GPU used for this research.

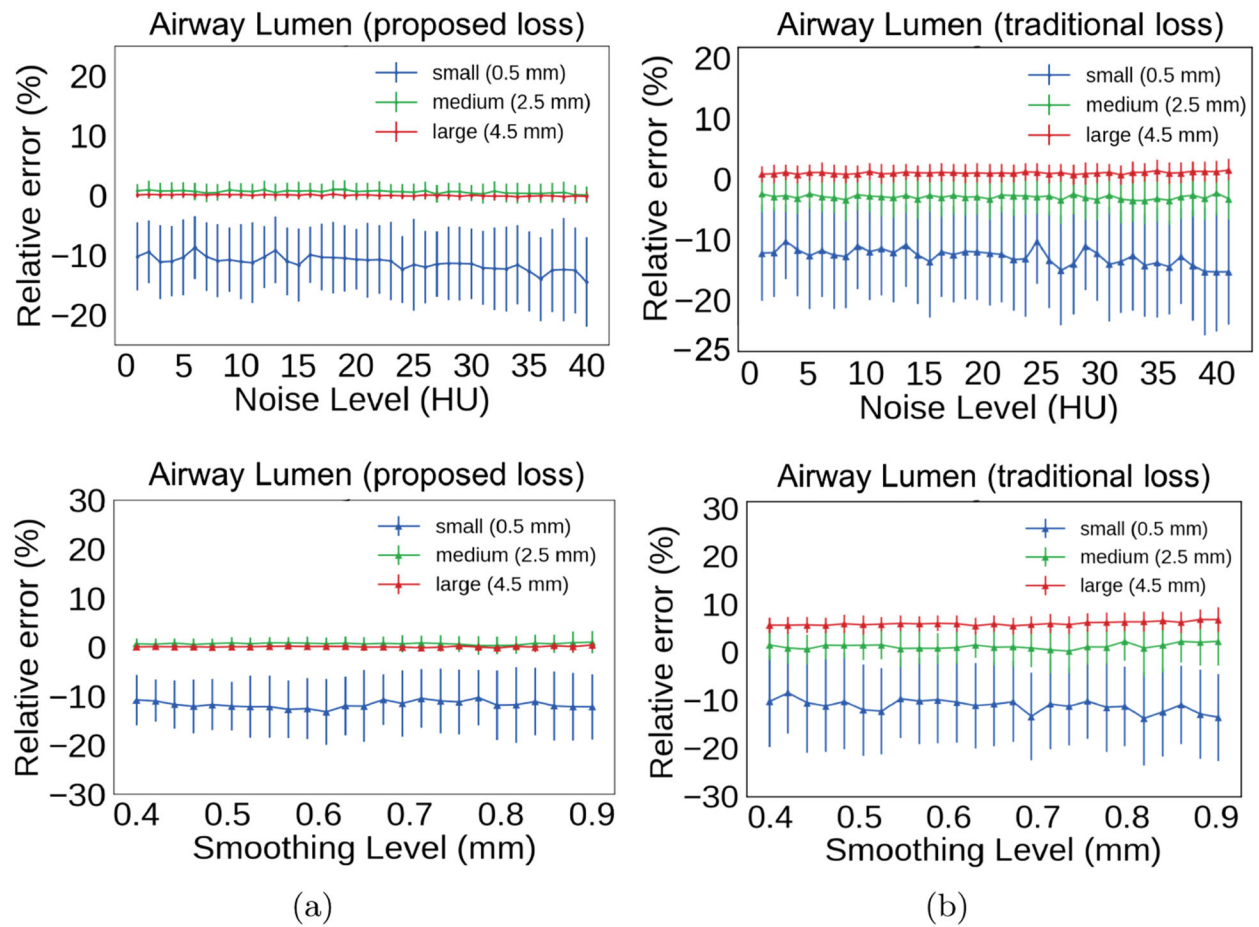
## References

1. Antoniou A, Storkey A, Edwards H: Data augmentation generative adversarial networks. arXiv preprint arXiv:1711.04340 (2017)
2. Cubuk ED, Zoph B, Mane D, Vasudevan V, Le QV: Autoaugment: learning augmentation policies from data. arXiv preprint arXiv:1805.09501 (2018)
3. Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H: GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* 321, 321–331 (2018)
4. Goodfellow I, et al.: Generative adversarial nets. In: *Advances in Neural Information Processing Systems*, pp. 2672–2680 (2014)
5. McCann MT, et al.: Convolutional neural networks for inverse problems in imaging: a review. *IEEE SPM* 34(6), 85–95 (2017)
6. Nardelli P, et al.: Accurate measurement of airway morphology on chest CT images In: Stoyanov D, et al. (eds.) *RAMBO/BIA/TIA –2018. LNCS*, vol. 11040, pp. 335–347. Springer, Cham (2018). 10.1007/978-3-030-00946-534
7. Ronneberger O, Fischer P, Brox T: U-net: convolutional networks for biomedical image segmentation In: Navab N, Hornegger J, Wells WM, Frangi AF (eds.) *MICCAI 2015. LNCS*, vol. 9351, pp. 234–241. Springer, Cham (2015). 10.1007/978-3-319-24574-4\_28
8. Ross JC, et al.: Lung extraction, lobe segmentation and hierarchical region assessment for quantitative analysis on high resolution computed tomography images In: Yang G-Z, Hawkes D, Rueckert D, Noble A, Taylor C (eds.) *MICCAI 2009. LNCS*, vol. 5762, pp. 690–698. Springer, Heidelberg (2009). 10.1007/978-3-642-04271-384
9. Suzuki K: Overview of deep learning in medical imaging. *Radiol. Phys. Technol* 10(3), 257–273 (2017) [PubMed: 28689314]

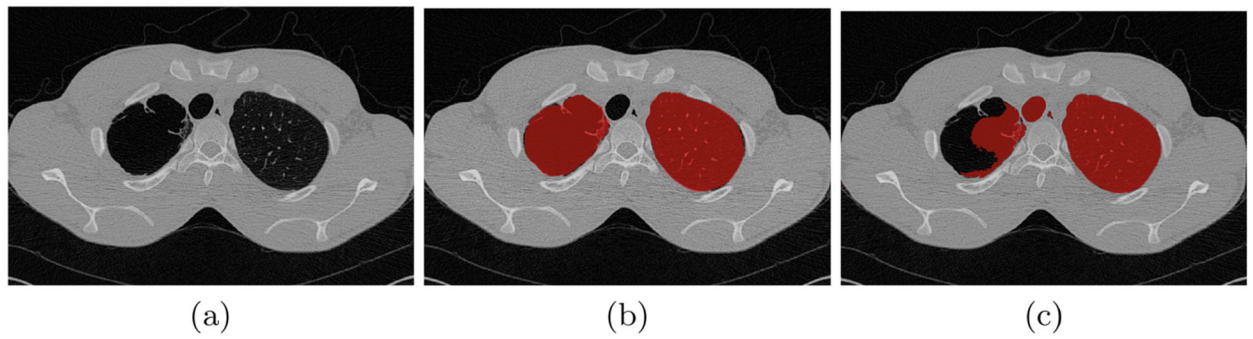
**Fig.1.**

Scheme of accuracy-precision loss. Accuracy is computed over all  $N \times M$  images (red square), while precision is calculated over the  $M$  image replicas (purple square).

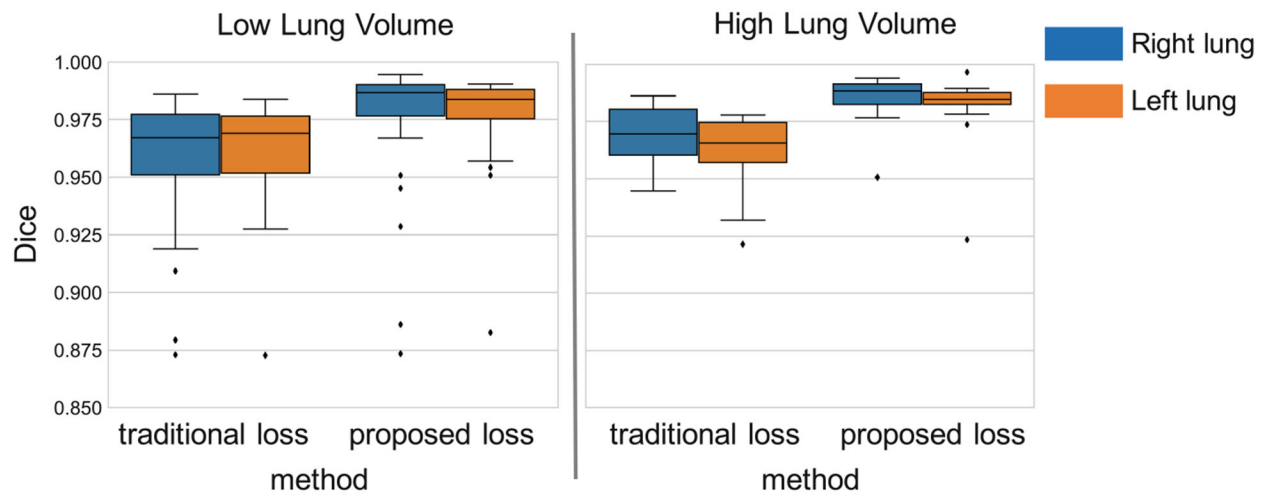


**Fig.2.**

Results obtained for airway lumen regression when varying the level of noise (first row) and smoothing (second row). In both cases, the RE (reported in %) obtained with proposed loss (a) is lower than that obtained when precision is not taken into account (b). Also, a precision-accuracy loss function greatly reduces the standard deviation for the RE.



**Fig.3.** Example of lung segmentation for case 23 from the LOLA11 dataset. A slice from the original CT (a), and the same slice overlaid with the segmentation obtained using (b) the proposed loss function and (c) a Dice loss function are presented.



**Fig.4.** Dice coefficient results obtained with the traditional and the proposed method when considering right and left lung independently and separating the cases by lung volume.