Temporal Convolution Networks for Real-Time Abdominal Fetal Aorta Analysis with Ultrasound

Nicoló Savioli[1(✉)], Silvia Visentin[2(✉)], Erich Cosmi[2(✉)], Enrico Grisan[1,3(✉)], Pablo Lamata[1(✉)], and Giovanni Montana[1,4(✉)]

[1] Department of Biomedical Engineering, Kings College London, London SE1 7EH, UK
{nicolo.l.savioli,enrico.grisan,pablo.lamata,giovanni.montana}@kcl.ac.uk
[2] Department of Woman and Child Health, University Hospital of Padova, Padua, Italy
{silvia.visentin.1,erich.cosmi}@unipd.it
[3] Department of Information Engineering, University of Padova, Padua, Italy
enrigri@dei.unipd.it
[4] WMG, University of Warwick, Coventry CV4 71AL, UK
g.montana@warwick.ac.uk

Abstract. The automatic analysis of ultrasound sequences can substantially improve the efficiency of clinical diagnosis. In this work we present our attempt to automate the challenging task of measuring the vascular diameter of the fetal abdominal aorta from ultrasound images. We propose a neural network architecture consisting of three blocks: a convolutional layer for the extraction of imaging features, a Convolution Gated Recurrent Unit (C-GRU) for enforcing the temporal coherence across video frames and exploiting the temporal redundancy of a signal, and a regularized loss function, called *CyclicLoss*, to impose our prior knowledge about the periodicity of the observed signal. We present experimental evidence suggesting that the proposed architecture can reach an accuracy substantially superior to previously proposed methods, providing an average reduction of the mean squared error from $0.31\,\mathrm{mm}^2$ (state-of-art) to $0.09\,\mathrm{mm}^2$, and a relative error reduction from 8.1% to 5.3%. The mean execution speed of the proposed approach of 289 frames per second makes it suitable for real time clinical use.

Keywords: Cardiac imaging · Diameter · Ultrasound
Convolutional networks · Fetal imaging · GRU · CyclicLoss

1 Introduction

Fetal ultrasound (US) imaging plays a fundamental role in the monitoring of fetal growth during pregnancy and in the measurement of the fetus well-being. Growth monitoring is becoming increasingly important since there is an

epidemiological evidence that abnormal birth weight is associated with an increased predisposition to diseases related to cardiovascular risk (such as diabetes, obesity, hypertension) in young and adults [1].

Among the possible biomarkers of adverse cardiovascular remodelling in fetuses and newborns, the most promising ones are the Intima-Media Thickness (IMT) and the stiffness of the abdominal aorta by means of ultrasound examination. Obtaining reliable measurements is critically based on the accurate estimation of the diameter of the aorta over time. However, the poor signal to noise ratio of US data and the fetal movement makes the acquisition of a clear and stable US video challenging. Moreover, the measurements rely either on visual assessment at bed-side during patient examination, or on tedious, error-prone and operator-dependent review of the data and manual tracing at later time. Very few attempts towards automated assessment have been presented [2,3], all of which have computational requirements that prevent them to be used in real-time. As such, they have reduced appeal for the clinical use. In this paper we describe a method for automated measurement of the abdominal aortic diameter directly from fetal US videos. We propose a neural network architecture that is able to process US videos in real-time and leverage both the temporal redundancy of US videos and the quasi-periodicity of the aorta diameter.

The main contributions of the proposed method are as follows. First we show that a shallow CNN is able to learn imaging features and outperforms classical methods as level-set for fetal abdominal aorta diameter prediction. Second we add to the CNN a Convolution Gated Recurrent Unit (C-GRU) [15] for exploiting the temporal redundancy of the features extracted by CNN from the US video sequence. Finally, we add a new penalty term to the loss function used to train the CNN to exploit periodic variations.

2   Related Work

The interest for measuring the diameter and intima-media thickness (IMT) of major vessels has stemmed from its importance as biomarker of hypertension damage and atherosclerosis in adults. Typically, the IMT is assessed on the carotid artery by identifying its lumen and the different layers of its wall on high resolution US images. The improvements provided by the design of semi-automatic and automatic methods based mainly on the image intensity profile, distribution and gradients analysis, and more recently on active contours. For a comprehensive review of these classical methods we refer the reader to [4] and [5]. In the prenatal setting, the lower image quality, due to the need of imaging deeper in the mother's womb and by the movement of the fetus, makes the measurement of the IMT biomarker, although measured on the abdominal aorta, challenging.

Methods that proved successful for adult carotid image analysis do not perform well on such data, for which only a handful of methods (semi-automatic or automatic) have been proposed, making use of classical tracing methods and mixture of Gaussian modelling of blood-lumen and media-adventitia interfaces [2],

or on level sets segmentation with additional regularizing terms linked to the specific task [3]. However, their sensitivity to the image quality and lengthy computation prevented an easy use in the clinical routine.

Deep learning approaches have outperformed classical methods in many medical tasks [8]. The first attempt in using a CNN, for the measurement of carotid IMT has been made only recently [9]. In this work, two separate CNNs are used to localize a region of interest and then segment it to obtain the lumen-intima and media-adventitia regions. Further classical post-processing steps are then used to extract the boundaries from the CNN based segmentation. The method assumes the presence of strong and stable gradients across the vessel walls, and extract from the US sequence only the frames related to the same cardiac phase, obtained by a concomitant ECG signal.

However, the exploitation of temporal redundancy on US sequences was shown to be a solution for improving overall detection results of the fetal heart [11], where the use of a CNN coupled with a recurrent neural network (RNN) is strategic. Other works, propose similar approach in order to detect the presence of standard planes from prenatal US data using CNN with Long-Short Term Memory (LSTM) [10].

3   Datasets

This study makes use of a dataset consisting of 25 ultrasound video sequences acquired during routine third-trimester pregnancy check-up at the Department of Woman and Child Health of the University Hospital of Padova (Italy). The local ethical committee approved the study and all patients gave written informed consent.

Fetal US data were acquired using a US machine (Voluson E8, GE) equipped with a 5 MHz linear array transducer, according to the guidelines in [6,7], using a 70° FOV, image dimension 720×960 pixels, a variable resolution between 0.03 and 0.1 mm and a mean frame rate of 47 fps. Gain settings were tuned to enhance the visual quality and contrast during the examination. The length of the video is between 2 s and 15 s, ensuring that at least one full cardiac cycle is imaged.

After the examination, the video of each patient was reviewed and a relevant video segment was selected for semi-automatic annotation considering its visual quality and length: all frames of the segment were processed with the algorithm described in [2] and then the diameters of all frames in the segments were manually reviewed and corrected. The length of the selected segments varied between 21 frames 0.5 s and 126 frames 2.5 s. The 25 annotated segments in the dataset were then randomly divided into training (60% of the segments), validation (20%) and testing (20%) sets. In order to keep the computational and memory requirements low, each frame was cropped to have a square aspect ratio and then resized to 128×128 pixels. The data supporting this research are [openly available].

4   Network Architecture

Our output is the predicted value $\hat{y}[t]$ of the diameter of the abdominal aorta at each time point. Our proposed deep learning solution consists of three main components (see Fig. 1): a Convolutional Neural Network (CNN) that captures the salient characteristics from ultrasound input images; a Convolution Gated Recurrent Unit (C-GRU) [15] exploits the temporal coherence through the sequence; and a regularized loss function, called *CyclicLoss*, that exploits the redundancy between adjacent cardiac cycles.

Our input consists of a set of sequences whereby each sequence $S = [s[1], ..., s[K]]$ has dimension $N \times M$ pixels at time t, with $t \in \{1, \ldots, K\}$. At each time point t, the CNN extracts the feature maps $x[t]$ of dimensions $D \times N_x \times M_x$, where D is the number of maps, and N_x and M_x are their in-plane pixel dimensions, that depend on the extent of dimensionality reduction obtained by the CNN through its pooling operators.

The feature maps are then processed by a C-GRU layer [15]. The C-GRU combines the current feature maps $x[t]$ with an encoded representation $h[t-1]$ of the feature maps $\{x[1], \ldots, x[t-1]\}$ extracted at previous time points of the sequence to obtain an updated encoded representation $h[t]$, the *current state*, at time t: this allows to exploit the temporal coherence in the data. The $h[t]$ of the C-GRU layer is obtained by two specific gates designed to control the information inside the unit: a reset gate, $r[t]$, and an update gate, $z[t]$, defined as follows:

$$r[t] = \sigma(W_{hr} * h[t-1] + W_{xr} * x[t] + b_r) \tag{1}$$

$$z[t] = \sigma(W_{hz} * h[t-1] + W_{xz} * x[t] + b_z) \tag{2}$$

Where, $\sigma()$ is the sigmoid function, $W.$ are recurrent weights matrices whose first subscript letter refers to the input of the convolution operator (either the feature maps $x[t]$ or the state $h[t-1]$), and whose second subscript letter refers to the gate (reset r or update z). All this matrices, have a dimension of $D \times 3 \times 3$ and $b.$ is a bias vector. In this notation, $*$ defines the convolution operation. The current state is then obtained as:

$$h[t] = (1 - z[t]) \odot h[t-1] + z[t] \odot \tanh(W_h * (r[t] \odot h_{t-1}) + W_x * x[t] + b). \tag{3}$$

Where $\odot$ denotes the dot product and W_h and W_x are recurrent weight matrices for $h[t-1]$ and $x[t]$, used to balance the new information represented by the feature maps $x[t]$ derived by the current input data $s[t]$ with the information obtained observing previous data $s[1], \ldots, s[t-1]$. On the one hand, $h[t]$ is then passed on for updating the state $h[t+1]$ at the next time point, and on the other is flatten and fed into the last part of the network, built by Fully Connected (FC) layers progressively reducing the input vector to a scalar output that represent the current diameter estimate $\hat{y}[t]$.
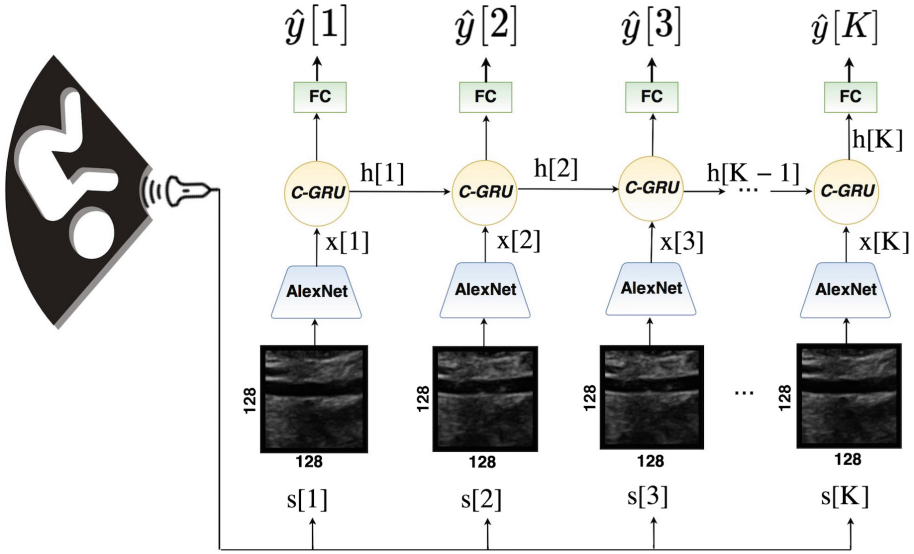
Fig. 1. The deep-learning architecture proposed for abdominal diameter aorta prediction. The blue blocks represent the features extraction through a CNN (AlexNet) which takes in input a US sequence S, and provides for each frame $s[t]$ a features map $x[t]$ that is passed to Convolution Gated Recurrent Units (C-GRU) (yellow circle) that encodes and combines the information from different time points to exploit the temporal coherence. The fully connected block (FC, in green), takes as input the current encoded state $h[t]$ as features to estimate the aorta diameter $\hat{y}[t]$. (Color figure online)

4.1   CyclicLoss

Under the assumption that the pulsatility of the aorta follows a periodic pattern with the cardiac cycle, the diameter of the vessel at corresponding instants of the cardiac cycle should ideally be equal. Assuming a known cardiac period T_{period}, we propose to add a regularization term to the loss function used to train the network as to penalize large differences of the diameter values that are estimated at time points that are one cardiac period apart.

We call this regularization term *CyclicLoss* (CL), computed as L_2 norm between pairs of predictions at the same point of the heart cycle and from adjacent cycles:

$$CL = \sqrt{\sum_{n=1}^{N_{cycles}} \sum_{t=0}^{T_{period}} \| \, \hat{y}[t + (n-1)T_{period}] - \hat{y}[t + nT_{period}] \, \|_2} \qquad (4)$$

The T_{period} is the period of the cardiac cycle, while N_{cycles} is the number of integer cycles present in the sequence and $\hat{y}[t]$ is the estimated diameter at time t. Notably, the T_{period} is determined through a peak detection algorithm on $y[t]$, and the average of all peak-to-peak detection distances define its value.

While the N_{cycles} is the number of cycles present, calculated as the total length of the $y[t]$ signal divided by T_{period}.

The loss to be minimized is therefore a combination of the classical mean squared error (MSE) with the CL, and the balance between the two is controlled by a constant λ:

$$Loss = MSE + \lambda \cdot CL = \frac{1}{K} \sum_{t=1}^{K} (y[t] - \hat{y}[t])^2 + \lambda \cdot CL \qquad (5)$$

where $y[t]$ is the target diameter at time point t. It is worth noting that the knowledge of the period of the cardiac cycle is needed only during training phase. Whereas, during the test phase, on unknown image sequence, the trained network provide its estimate blind of the periodicity of the specific sequence under analysis.
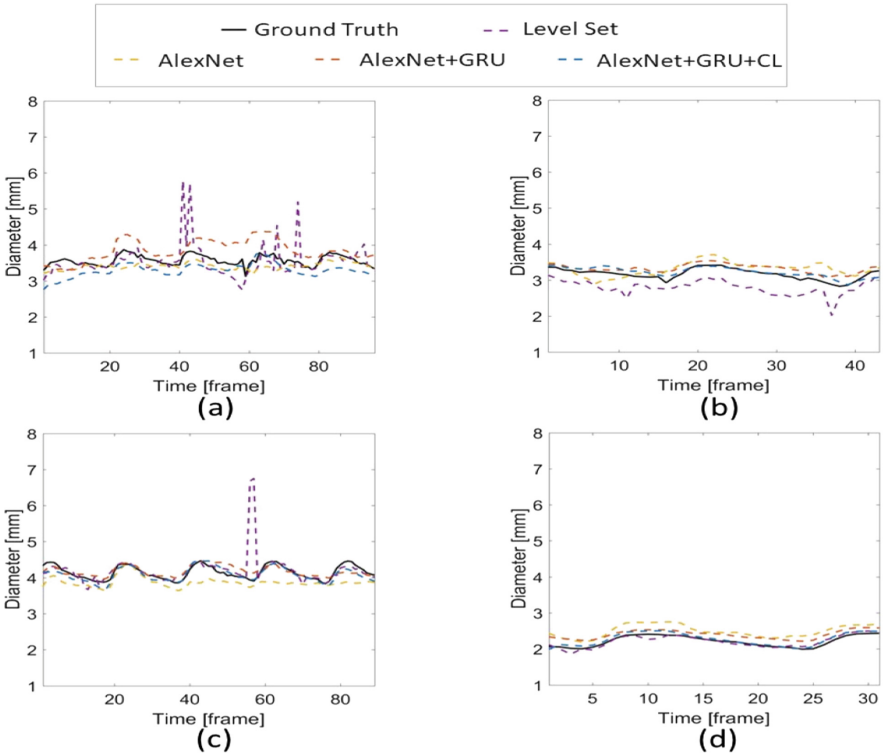

Fig. 2. Each panel (a–c) shows the estimation of the aortic diameter at each frame of fetal ultrasound videos in the test set, using the level set method (dashed purple line), the naive architecture using AlexNet (dashed orange line), the AlexNet+C-GRU (dashed red line), and AlexNet+C-GRU trained with the *CyclicLoss* (dashed blue line). The ground truth (solid black line) is reported for comparison. Panels (a, c) show the results on long sequences where more than 3 cardiac cycles are imaged, whereas panels (b, d) show the results on short sequences where only 1 or two cycles are available. (Color figure online)

4.2   Implementation Details

For our experiments, we chose AlexNet [12] as a feature extractor for its simplicity. It has five hidden layers with 11×11 kernels size in the first layer, 5×5 in the second and 3×3 in the last three layers; it is well suited to the low image contrast and diffuse edges characteristic of US sequences. Each network input for the training is a sequence of $K = 125$ ultrasound frames with $N = M = 128$ pixels, AlexNet provides feature maps of dimension $D \times N \times M = 256 \times 13 \times 13$, and the final output $\hat{y}[t]$ is the estimate abdominal aorta diameter value at each frame.

The loss function is optimised with the Adam algorithm [16] that is a first-order gradient-based technique. The learning rate used is $1e^{-4}$ with 2125 iterations (calculated as number of patients × number of ultrasound sequences) for 100 epochs. In order to improve generalization, data augmentation of the input with a vertical and horizontal random flip is used at each iteration. The λ constant used during training with *CyclicLoss* takes the value of $1e^{-6}$.

5   Experiments

The proposed architecture is compared with the currently adopted approach in Sect. 4. This method provides fully-automated measurements in lumen identification on prenatal US images of the abdominal aorta [3] based on edge-based level set. In order to understand the behaviour of different features extraction methods, we have also explored the performance of new deeper network architectures whereby AlexNet was replaced it by InceptionV4 [13] and DenseNets 121 [14].

Table 1. The table show the mean (standard deviation) of MSE and RE error for all the comparison models. The combination of C-GRU and the *CyclicLoss* with AlexNet yields the best performance. Adding recurrent units to any CNN architecture improves its performance; however deeper networks as InceptionV4 and DenseNets do not show any particular benefits with respect to the simpler AlexNet. Notably, we also consider the p-value for multiple models comparison with the propose network AlexNet+C-GRU+CL, in this case the significant level should be 0.05/7 using the Bonferroni correction [17].

Methods	MSE [mm^2]	RE [%]	p-value
AlexNet	0.29(0.09)	8.67(10)	1.01e−12
AlexNet+C-GRU	0.093(0.191)	6.11(5.22)	1.21e−05
AlexNet+C-GRU+CL	**0.085(0.17)**	**5.23(4.91)**	"-"
DenseNet121	0.31(0.56)	9.55(8.52)	6.00e−13
DenseNet121+C-GRU	0.13(0.21)	7.72(5.46)	7.78e−12
InceptionV4	6.81(14)	50.4(39.5)	6.81e−12
InceptionV4+C-GRU	0.76(1.08)	16.3(9.83)	2.89e−48
Level-set	0.31(0.80)	8.13(9.39)	1.9e−04

The performance of each method was evaluated both with respect to the mean squared error (MSE) and to the mean absolute relative error (RE); all values are reported in Table 1 in terms of average and standard deviation across the test set.

In order to provide a visual assessment of the performance, representative estimations on four sequences of the test set are shown in Fig. 2. The naive architecture relying on a standard loss and its C-GRU version are incapable to capture the periodicity of the diameter estimation. The problem is mitigated by adding the *CyclicLoss* regularization on MSE. This is quantitatively shown in Table 1, where the use of this loss further decreases the MSE from $0.093\,\mathrm{mm}^2$ to $0.085\,\mathrm{mm}^2$, and the relative error of from 6.11% to 5.23%.

Strikingly, we observed that deeper networks are not able to outperform AlexNet on this dataset. Their limitation may be due to over-fitting. Nevertheless, the use of C-GRU greatly improve the performance of both networks both in terms of MSE and of RE. Further, we also performed a non-parametric test (Kolmogorov-Smirnov test) to check if the best model was statistically different compared to the others.

The results obtained with the complete model AlexNet+C-GRU+CL are indeed significantly different from all others ($p < 0.05$) also, when the significant level is adjusted for multiple comparison applying the Bonferroni correction [17,18].

6   Discussion and Conclusion

The deep learning (DL) architecture proposed shows excellent performance compared to traditional image analysis methods, both in accuracy and efficiency. This improvement is achieved through a combination of a shallow CNN and the exploitation of the temporal and cyclic coherence. Our results seem to indicate that a shallow CNNs perform better than deeper CNNs such as DenseNet 121 and InceptionV4; this might be due to the small dimension of the data set, a common issue in the medical settings when requiring manual annotations of the data.

6.1   The *CyclicLoss* Benefits

The exploitation of temporal coherence is what pushes the performance of the DL solution beyond current image analysis methods, reducing the MSE from $0.29\,\mathrm{mm}^2$ (naive architecture) to $0.09\,\mathrm{mm}^2$ with the addition of the C-GRU. The *CyclicLoss* is an efficient way to guide the training of the DL solution in case of data showing some periodicity, as in cardiovascular imaging. Please note that the knowledge of the signal period is only required by the network during training, and as such it does not bring additional requirements on the input data for real clinical application. We argue that the *CyclicLoss* is making the network learn to expect a periodic input and provide some periodicity in the output sequence.

6.2   Limitations and Future Works

A drawback of this work is that it assumes the presence of the vessel in the current field of view. Further research is thus required to evaluate how well the solution adapts to the scenario of lack of cyclic consistency, when the vessel of interest can move in and out of the field of view during the acquisition, and to investigate the possibility of a concurrent estimation of the cardiac cycle and vessel diameter. Finally, the C-GRU used in our architecture, has two particular advantages compared to previous approaches [10,11]: first, it is not subject to the vanishing gradient problem as the RNN, allowing to train from long sequences of data. Second, it has less computational cost compared to the LSTM, and that makes it suitable for real time video application.

References

1. Visentin, S., Grumolato, F., Nardelli, G.B., Di Camillo, B., Grisan, E., Cosmi, E.: Early origins of adult disease: low birth weight and vascular remodeling. Atherosclerosis **237**(2), 391–399 (2014)
2. Veronese, E., Tarroni, G., Visentin, S., Cosmi, E., Linguraru, M.G., Grisan, E.: Estimation of prenatal aorta intima-media thickness from ultrasound examination. Phys. Med. Biol. **59**(21), 6355–6371 (2014)
3. Tarroni, G., Visentin, S., Cosmi, E., Grisan, E.: Fully-automated identification and segmentation of aortic lumen from fetal ultrasound images. In: IEEE EMBC, pp. 153–156 (2015)
4. Molinari, F., Zeng, G., Suri, J.S.: A state of the art review on intimamedia thickness (IMT) measurement and wall segmentation techniques for carotid ultrasound. Comp. Meth. Prog. Biomed. **100**(3), 201–221 (2010)
5. Loizou, C.P.: A review of ultrasound common carotid artery image and video segmentation techniques. Med. Biol. Eng. Comp **52**(12), 1073–1093 (2014)
6. Cosmi, E., Visentin, S., Fanelli, T., Mautone, A.J, Zanardo, V.: Aortic intima media thickness in fetuses and children with intrauterine growth restriction. Obs. Gyn. **114**, 1109–1114 (2009)
7. Skilton, M.R., Evans, N., Griffiths, K.A., Harmer, J.A., Celermajer, D.S.: Aortic wall thickness in newborns with intrauterine growth restriction. Lancet **365**, 1484–14846 (2005)
8. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
9. Shin, J.Y., Tajbakhsh, N., Hurst, R.T., Kendall, C.B., Liang, J.: Automating carotid intima-media thickness video interpretation with convolutional neural networks. In: IEEE CVPR Conference, pp. 2526–2535 (2016)
10. Chen, H., et al.: Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 507–514. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_62

11. Huang, W., Bridge, C.P., Noble, J.A., Zisserman, A.: Temporal HeartNet: towards human-level automatic analysis of fetal cardiac screening video. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10434, pp. 341–349. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66185-8_39
12. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: NIPS 2012, pp. 1097–1105 (2012)
13. Szegedy, C., Ioffe, S., Vanhoucke, V.: Inception-v4, inception-ResNet and the impact of residual connections on learning. In: AAAI 2017, pp. 4278–4284 (2017)
14. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: IEEE CVPR Conference, pp. 2261–2269 (2017)
15. Siam, M., Valipour, A., Jägersand, M., Ray, N.: Convolutional gated recurrent networks for video segmentation. In: IEEE ICIP Conference, pp. 3090–3094 (2017)
16. Kingma, D.P., Ba, L.J.: Adam: a method for stochastic optimization. In: 3rd International Conference for Learning Representations (2015)
17. Bonferroni, C.E.: Teoria statistica delle classi e calcolo delle probabilit. Pubblicazioni del Regio Istituto Superiore di Scienze Economiche e Commerciali di Firenze (1936)
18. Dunn, O.J.: Multiple comparisons among means. J. Am. Stat. Assoc. **56**(293), 52–64 (1961)