MuTGAN: Simultaneous Segmentation and Quantification of Myocardial Infarction without Contrast Agents via Joint Adversarial Learning

Chenchu Xu[1], Lei Xu[2], Gary Brahm[1], Heye Zhang[3⋆], Shuo Li[1⋆]

[1] Western university, London ON, Canada
[2] Beijing AnZhen Hospital, Beijing, China
[3] Shenzhen Institutes of Advanced Technology,
Chinese Academy of Sciences, Shenzhen, China

Abstract. Simultaneous segmentation and full quantification (estimation of all diagnostic indices) of the myocardial infarction (MI) area are crucial for early diagnosis and surgical planning. Current clinical methods still suffer from high-risk, non-reproducibility and time-consumption issues. In this study, the multitask generative adversarial networks (MuTGAN) is proposed as a contrast-free, stable and automatic clinical tool to segment and quantify MIs simultaneously. MuTGAN consists of generator and discriminator modules and is implemented by three seamless connected networks: spatio-temporal feature extraction network comprehensively learns the morphology and kinematic abnormalities of the left ventricle through a novel three-dimensional successive convolution; joint feature learning network learns the complementarity between segmentation and quantification through innovative inter- and intra-skip connection; task relatedness network learns the intrinsic pattern between tasks to increase the accuracy of estimations through creatively utilized adversarial learning. MuTGAN minimizes a generalized divergence to directly optimize the distribution of estimations by using the competition process, which achieves pixel segmentation and full quantification of MIs. Our proposed method yielded a pixel classification accuracy of 96.46%, and the mean absolute error of the MI centroid was 0.977 *mm*, from 140 clinical subjects. These results indicate the potential of our proposed method in aiding standardized MI assessments.

1 Introduction

Simultaneous segmentation and quantification (*including pixel segmentation and full quantification of all indices such as the infarct size, segment percentage, perimeter, centroid, major axis length, minor axis length and orientation*) of a myocardial infarction (MI) are crucial to clinical treatment of the MI [1]. It segments MI to predict the recovery of dysfunctional segments in chronic ischemic heart diseases or to select therapeutic options; it estimates all indices that indicate the presence, location, and transmurality of acute and chronic MI. The combination can obtain all information required for a thorough understanding

[⋆] Corresponding Author: Dr. Shuo Li (slishuo@gmail.com) and Dr. Heye Zhang (hy.zhang@siat.ac.cn)

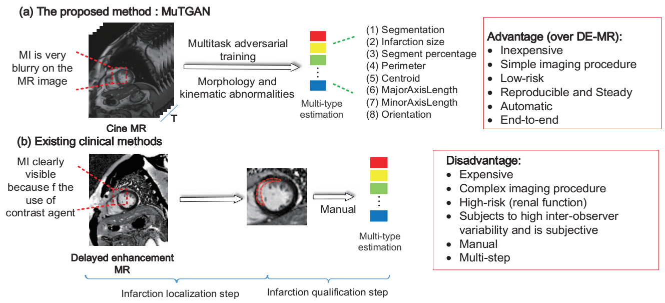

(a) The proposed method : MuTGAN
MI is very blurry on the MR image
Cine MR
T
Multitask adversarial training
Morphology and kinematic abnormalities
Multi-type estimation
(1) Segmentation
(2) Infarction size
(3) Segment percentage
(4) Perimeter
(5) Centroid
(6) MajorAxisLength
(7) MinorAxisLength
(8) Orientation
Advantage (over DE-MR):
Inexpensive
Simple imaging procedure
Low-risk
Reproducible and Steady
Automatic
End-to-end
(b) Existing clinical methods
MI clearly visible because f the use of contrast agent
Delayed enhancement MR
Manual
Multi-type estimation
Disadvantage:
Expensive
Complex imaging procedure
High-risk (renal function)
Subjects to high inter-observer variability and is subjective
Manual
Multi-step
Infarction localization step
Infarction qualification step

Fig. 1: (a) MuTGAN can accurately segment and quantify an infarction in one step without contrast agents compared to (b) the high-risk, non-reproducible and time-consuming current clinical methods.

of the extent of the MI to prevent further heart failure [2].

Current clinical methods still suffer from high-risk, non-reproducibility and time-consumption issues [3]. The clinical standards include two steps: 1) delayed enhancement (DE) imaging by using magnetic resonance (MR) with gadolinium contrast agents and 2) manual segmentation and quantification of all indices of the MI from DE -MR image [2]. The high risk comes from the use of contrast agents, which are fatal with regard to the kidney disease that accompanies more than 20% of MI patients [3]. The non-reproducibility comes from the manual process that is subject to high inter-observer variability and is subjective. The time-consumption factor comes from the imaging process itself that requires multiple imaging techniques and a two-step process. Therefore, there is an urgent clinical desire to obtain a non-contrast agent, stable and automated clinical tool. However, while it is widely believed that an MI can be identified and localized vaguely without a contrast agent through the detection of morphological and kinematic abnormalities of the left ventricle (LV) directly from blurry cine MR [4], it is still challenging to build a unified model for segmenting and quantifying the MI simultaneously: 1) Effective learning of the relationship between segmentation and quantification. Segmentation and quantification as two related tasks sharing the same factors can share the information during the learning process to produce a beneficial interaction [5]. However, it is difficult to uniformly learn this beneficial interaction because of huge differences in the dimensions and distribution between the two. 2) Comprehensive learning of the spatio-temporal information inside of and between images. Extracting spatio-temporal information from the myocardium and surrounding tissues in sequential images is highly effective for building the intrinsic representation of the MI [6]. However, it is still difficult to systematically learn the asymmetry of the spatial and temporal motion over different time steps from 2D+T image sequences [7]. 3) Efficient leveraging of the relationship between quantification indices. Accurate full quantification is highly dependent on the ability to leverage the important relationship that exists between the different indices to directly optimize the estimations [8]. However, it is difficult to train this relationship properly to reduce the distribution error of the estimation due to the different locally optimal and probability distributions of the different indices [9].

In this study, the multitask generative adversarial networks (MuTGAN) is proposed for joint segmentation and quantification of MI directly from cine MR without contrast agents. MuTGAN formulates the segmentation and seven quantification indices into eight tasks that can be estimated simultaneously under combined multitask and adversarial learning. To accomplish this, a novel three dimensional (3D) successive convolutional framework that takes spatial correlation over different time steps into consideration is proposed for extracting the comprehensive spatio-temporal feature of MI from 2D+T images. An innovative joint learning architecture as multitask learning that fuses the feature maps of different level layers is proposed for achieving a reciprocal representation that has a beneficial interaction between segmentation and quantification. A creative task relatedness adversarial learning that models the substantial pattern between tasks is proposed for exploiting the inductive bias of the task's relevance to approximate the estimation distribution to the real data. In the end, our MuTGAN not only simultaneously segments and quantifies MI for the first time but also effectively exploits the trait of multitask to improve itself.

2 Methodology

MuTGAN was implemented by two competing modules: the generator (Sect. 2.2) and discriminator (Sect. 2.3), as shown in Fig. 2. The two modules interact with each other and consist of three seamlessly connected networks. The generator first builds the spatio-temporal feature extraction network based on novelty stacks of 3D convolution (Conv) and ConvLSTMs [10] to learn a comprehensive representation of the 2D+T data through successive convolution over different time steps in replacing the pooling layer; it then builds joint feature learning network based on new skip architecture [11] to share the representation of the segmentation and quantification by multiple skip connections between inter- and intra-networks. The discriminator builds task relatedness network based on bidirectional (Bi) -LSTMs [12] to creatively use adversarial training that takes the complete contextual pattern between tasks as a criterion for measuring the accuracy of estimations.

2.1 MuTGAN Formulation

The output of MuTGAN is multitask result $Y \in (y_1, y_2, ..., y_n, n = 8)$ including one segmentation task y_1 and seven quantification tasks $y_2, ..., y_n$. The objective of MuTGAN is to simultaneously estimate Y from cine MR, which consists of $2D+T$ image sequences $X \in (x_1, x_2, ..., x_T, \mathbb{R}^{H\times W\times T})$, where H and W are the height and width of each temporal frames respectively ($H = W = 64$), and T is the temporal step, $T = 25$. X can be considered as special 3D data ($H \times W \times T$). Given the discriminator and generator parameters θ_D and θ_G, each updated by minimizing the losses $\mathcal{L}_G$ and $\mathcal{L}_D$, MuTGAN can be express as:

$$\begin{cases} \mathcal{L}_D = \mathcal{L}(d(Y)) - \lambda\mathcal{L}(d(g(X))) & \text{for } \theta_D \\ \mathcal{L}_G = \mathcal{L}(g(X)) & \text{for } \theta_G \end{cases} \tag{1}$$

Where $g(.)$ and $d(.)$ are the generator and discriminator function, and $\lambda \in [0, 1]$.

2.2 Generator

The generator module uses spatio-temporal feature extraction network and joint feature learning network to generate candidate multitask estimations directly

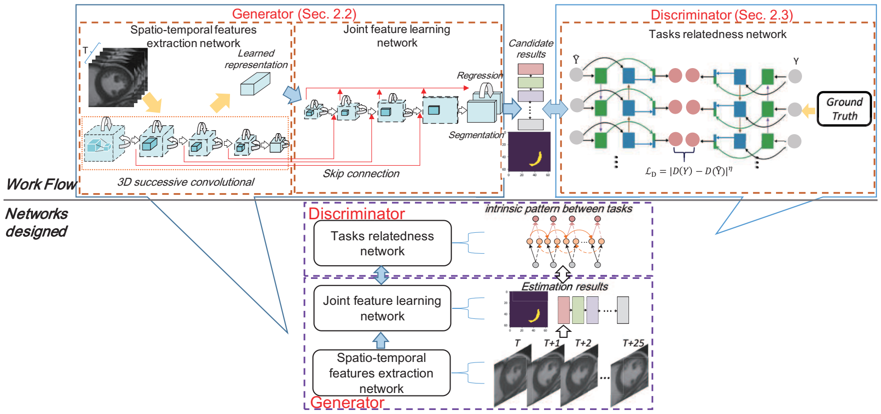

Generator (Sec. 2.2)
Spatio-temporal features extraction network
Learned representation
T
3D successive convolutional
Joint feature learning network
Regression
Segmentation
Skip connection
Candidate results
Discriminator (Sec. 2.3)
Tasks relatedness network
Ŷ
Y
Ground Truth
$\mathcal{L}_D = |D(Y) - D(\hat{Y})|^\eta$
Work Flow
Networks designed
Discriminator
Tasks relatedness network
intrinsic pattern between tasks
Joint feature learning network
Estimation results
Spatio-temporal features extraction network
Generator
T
T+1
T+2
T+25

Fig. 2: The architecture of MuTGAN: the generator combines spatio-temporal feature extraction network and joint feature learning network for multitask learning; the discriminator uses task relatedness network for adversarial learning.

from cine MR.

Spatio-temporal features extraction network (SFEN). The network innovatively stacks ConvLSTMs and 3DConvs for accurate extraction of the left ventricular morphology and kinematic abnormalities. ConvLSTMs and 3DConvs are both effective tools for learning spatial and temporal information from 2D+T data and are usually used separately. In our work, for a better handle on the asymmetry distortion that is a unique myocardial motion pattern caused by MI, special integration of ConvLSTMs and 3DConvs is used to consider spatial correlation over different time steps. The benefit of this integration is two-fold: first, by using successive Convs instead of the pooling layers, temporal and spatial information can be extracted simultaneously and equally. Second, multiple temporal ranges with different spatial scales can be extracted by adjusting the size of the Conv kernel over different layers. ConvLSTMs uses its internal memory to process 2D+T images, which creates an internal state that can discover the dynamic temporal behavior between frames and allows for persistence [10]. Given that i_t, f_t, $\tilde{c}_t$, and o_t represent the input, forget, cell, and output gates, respectively, the ConvLSTMs are:

$$\begin{cases} i_t &= \sigma(x_t * w_{xi} + h_{t-1} * w_{hi} + w_{i\text{bias}}) \\ f_t &= \sigma(x_t * w_{xf} + h_{t-1} * w_{hf} + w_{f\text{bias}}) \\ \tilde{c}_t &= \tanh(x_t * w_{x\tilde{c}} + h_{t-1} * w_{h\tilde{c}} + w_{\tilde{c}\text{bias}}) \\ c_t &= \tilde{c}_t \odot i_t + c_{t-1} \odot f_t \\ o_t &= \sigma(x_t * w_{xo} + h_{t-1} * w_{ho} + w_{o\text{bias}}) \\ h_t &= o_t \odot \tanh(c_t) \end{cases} \tag{2}$$

Where c and h are the memory and output activations; σ and tanh are the sigmoid and hyperbolic tangent non-linearities; $*$ represents the convolution operation, and $\odot$ the Hadamard product. 3DConvs are inserted between each ConvLSTMs, which stride 2×2×3 downsampling to reduce the resolution both spatially and temporally. In the last of this networks (Conv13), each X maps into a fixed-length vector $\mathcal{F}_{Conv13} = f_{encoding}(X) \approx \arg\min_{H} p((H - X) \mid x_1, \ldots, x_t), H \in (h_1, ..., h_t), \mathcal{F}_{Conv13} \in \mathbb{R}^{4\times4\times512}$.

Joint feature learning network. A new multiple skip connection is used to forcibly connect the segmentation and quantification, and joint learning beneficial interactions to promote the learning mutually. This network combines upsampling and the same successive Convs as SFEN uses the skip architecture to integrate SFEN and learn a shared representation of the segmentation and quantification. The skip architecture that fused feature maps of different layers to avoid spatial information loss and gradient dispersion has gained great success in U-Net and ResNet. In our work, for segmentation, the skip architecture symmetrically connects the SFEN and joint feature learning network ($Conv12 \rightarrow Conv18, Conv8 \rightarrow Conv22, Conv4 \rightarrow Conv26$) to contain the full context available in time series data and learn a more precise segmentation. For regression, the skip architecture combines the $\mathcal{F}_{Conv13}$ and each feature map before the filters change in the joint learning network to integrate coarse, high layer information with fine, low layer information to produce accurate and detailed quantification. Two parallel, fully connected layers generate candidate results $\widetilde{Y} \approx g_{predicting}(\mathcal{F}_{Conv13}) = g_{predicting}(\widetilde{y}_1, \ldots, \widetilde{y}_n \mid f_{encoding}(x_1, \ldots, x_t))$, including $\widetilde{y_1}$, a binary 64 × 64 image and seven indices $\widetilde{y_n}$.

Loss for generator training. Dice loss takes tackle class imbalance into consideration is employed for the segmentation and mean absolute error (MAE) for the quantification:

$$\mathcal{L}_G = \mathcal{L}_{dice}(g(X), y_1) + \sum_{i=2}^{n} \mathcal{L}_{mae}(g(X), y_i) = \frac{2|y_1 \cap g(X)|}{|y_1| + |g(X)|} + \sum_{i=2}^{n} |y_i - g(X)|^{\eta} \quad (3)$$

Where η is the target norm, $\eta \in \{1, 2\}$.

2.3 Discriminator

The discriminator creatively uses a task relatedness network to determine whether the candidate multitask estimations fit the ground truth and directly optimizes the candidate to improve MuTGAN performance based on the inherent pattern between tasks.

Task relatedness network. The network utilize Bi-LSTMs to learn the inherent pattern between tasks. LSTMs provides a general framework for learning this pattern. However, Bi-LSTMs consider more complete contextual relationships than LSTMs because of replacing each hidden sequence h^l with the forward and backward sequences $\overrightarrow{h^l}$ and $\overleftarrow{h^l}$, ensuring that every hidden layer receives input from both the forward and backward layers at the level below. In our work, Bi-LSTMs regulates task relatedness by gates structures, which remove or add information to the cell state by processing different tasks [12].

$$h_n^l = f_{LSTM}(W_{h^{l-1}h^l} h_n^{l-1} + W_{h^l h^l} h_{n-1}^l + b_h^l) \quad (4)$$

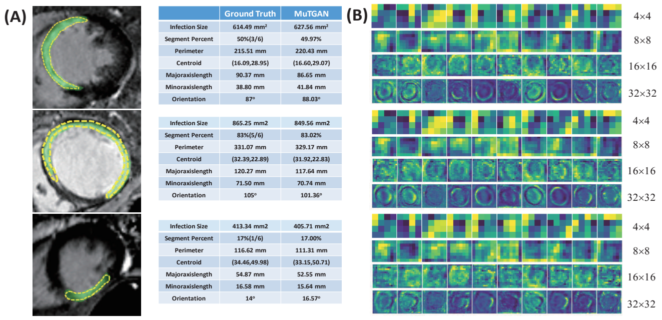
where the hidden vector sequences h_n^l are iteratively computed from $\tilde{Y} = \tilde{y}_1, \ldots, \tilde{y}_{\mathrm{n}}$ and all L ($l = 1$ to L) are layers in the stack. The network output is $\hat{Y}_l = W_{h^L \hat{y}} h_l^N + b_{\hat{y}}$

Loss for discriminator training. For efficiently converging GANs, the discriminator module is equivalent to minimizing the following MAE loss:

$$\mathcal{L}_D = \sum_{n=1}^{N} \mathcal{L}_{mae}(d(\tilde{y}_n), y_n) = \sum_{n=1}^{N} |y_n - d(\tilde{y}_n)|^{\eta} \quad (5)$$

3 Materials and Implementation Details

A total of 140 patients were retrospectively selected between May 2015 and May 2017 and completed cine and DE-MR imaging scans. MR imaging was performed



	Ground Truth	MuTGAN
Infection Size	614.49 mm²	627.56 mm²
Segment Percent	50%(3/6)	49.97%
Perimeter	215.51 mm	220.43 mm
Centroid	(16.09,28.95)	(16.60,29.07)
Majoraxislength	90.37 mm	86.65 mm
Minoraxislength	38.80 mm	41.84 mm
Orientation	87°	88.03°

Infection Size	865.25 mm2	849.56 mm2
Segment Percent	83%(5/6)	83.02%
Perimeter	331.07 mm	329.17 mm
Centroid	(32.39,22.89)	(31.92,22.83)
Majoraxislength	120.27 mm	117.64 mm
Minoraxislength	71.50 mm	70.74 mm
Orientation	105°	101.36°

Infection Size	413.34 mm2	405.71 mm2
Segment Percent	17%(1/6)	17.00%
Perimeter	116.62 mm	111.31 mm
Centroid	(34.46,49.98)	(33.15,50.71)
Majoraxislength	54.87 mm	52.55 mm
Minoraxislength	16.58 mm	15.64 mm
Orientation	14°	16.57°

Fig. 3: (A) The infarction areas segmented and quantified by our method (green zone) are consistent with the ground truth (yellow dotted line). (B) The different resolution feature maps indicate that MuTGAN effectively extracts spatio-temporal motion information of different myocardial regions at different time steps.

using a 3T MR system (MAGNETOM Verio, Siemens). SSFP cine images were acquired during repeated breath holds: TR 3.1 ms, TE 1.3 ms, FA 45°, FOV (276×340) mm^2, matrix 156×192, slice thickness 6 mm, and 25 cardiac phases. DE-MR imaging was performed in the same orientations and with the same slice thickness as cine imaging ten minutes after the intravenous injection of gadolinium (Magnevist, 0.2 m mol/kg): TR = 10.5 ms, TE = 5.4 ms, and FA = 30°. Two radiologists with more than 10 years of experience manually segmented and quantified the MI (syngo MR B17) in the DE-CMR images. If there was disagreement, a consensus between the two experts must be reached. In our experiments, a network with heart localization layers, described in [6], was used to automatically crop cine MR to 64 × 64 region-of-interest sequences including the LV. All experiments were assessed with a 10-fold cross-validation test.

4 Experiments and Results

MutGAN produces high performance with a pixel classification accuracy of 96.46%, the MAE of the centroid point is 0.97 mm with the ground truth obtained manually by human experts, which demonstrates this methods effectiveness for segmentation and quantification of MI.

Accurate MI Segmentation. The experiment's result shows that MuTGAN can accurately locate the MI, as shown in Fig. 3. We achieve an overall pixel classification accuracy of 96.46%, with a sensitivity of 91.82% and a specificity of 98.21%; the Dice coefficient is 90.27±0.05%; the ROCs and PRs curves are shown in Fig. 4. The ground truth and result are binary images; each pixel is assessed for infarction or normality (0 or 1).

Precise MI Quantification. MuTGAN can also obtain good quantification of the MI, as shown in Fig. 3 and Table 1. The MAE computed between the ground truth and our estimation of the infarction size is 22.311±18.39 mm^2, the segment percentage is 1.04±0.62%, the perimeter is 5.392±4.66 mm, the centroid is 0.977±0.78 mm, the major axis length is 2.303±1.88 mm, the minor axis length is 1.030±0.76 mm, and the orientation is 7.242±3.63°.

Advantage of GANs Architecture. Fig. 4 and Table 1 show that MuTGAN has better segmentation and quantification performance in comparison to those

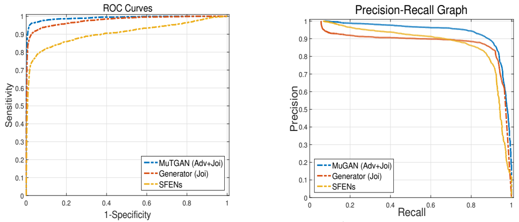


Fig. 4: ROCs and PRs demonstrate that MuTGAN can accurately segment infarct areas by combining joint learning (Joi) and adversarial learning (Adv).

frameworks because of its combined ability for joint learning (Joi) and adversarial learning (Adv). To evaluate this ability, MuTGAN (Joi+Adv), Generator (Joi) and separately estimated tasks by SFEN are implemented individually.

Table 1: MuTGAN works best in comparison with frameworks that do not utilize the joint learning and adversarial learning.

	MuTGAN	Generator	SFEN
Accuracy	**96.46%**	95.83%	94.13%
Sensitivity	**91.82%**	90.74%	81.86%
Specificity	**98.21%**	98.17%	96.64%
Dice (%)	**90.27%**	89.76%	83.13%
Infarct Size (mm^2)	**22.31**	30.91	54.91
segments percentage	**1.04%**	2.05%	1.05%
Perimeter (mm)	**5.39**	10.26	7.53
Centroid (mm)	**0.97**	3.87	1.42
Majoraxislength (mm)	**2.30**	4.83	7.65
Minoraxislength (mm)	**1.03**	3.26	5.34
Orientation (°)	**7.24**	8.51	8.47

Advantages of the Spatio-temporal Feature Extraction Network. Table 2 indicates that SFEN (3DConvs + ConvLSTMs) achieved better performance than all other the frameworks because of its learned spatial correlation over different time steps. To evaluate the performance of extraction, we replaced the 3DConvs + ConvLSTMs with 3DConvs, ConvLSTMs, 3DConvs + LSTMs, CNNs and LSTMs in our framework.

Table 2: MuTGAN works best in comparison with other frameworks.

	MuTGAN	ConvLSTMs	3DConvs	3DConvs+LSTM	CNNs	LSTMs
Accuracy	**96.46%**	96.31%	93.32%	94.48%	90.64%	90.63%
Dice	**0.90**	0.89	0.81	0.82	0.73	0.77
Infarct size	**22.31**	47.35	92.54	79.94	224.28	176.46

Comparison with Some Existing Methods. Table 3 demonstrates that MuTGAN achieved higher segmentation and quantification accuracy compared with existing classical methods in segmentation/quantification (Seq/Qua).

Table 3: MuTGAN realized segmentation and quantification of MI simultaneously without a contrast agent and yielded higher performance than some existing methods.

	Seg/Qua	Accuracy	Dice	Infarct size	Centroid
MuTGAN	**Seg and Qua**	**96.46%**	**90.27%**	**22.31**	**0.97**
Xu et al.[6]	Seg only	94.35%	89.87%	92.02∗	7.98∗
Popescu et al.[7]	Seg only	85.68%	74.45%	176.28∗	10.46∗
Bleton et al.[13]	Seg only	84.78	71.62	224.28∗	12.93∗

∗ The quantification result estimate from segmentation result.

5 Conclusions

Multitask generative adversarial networks have been proposed and, for the first time, used for simultaneous segmentation and quantification of MI without contrast agents. MuTGAN was conducted on 140 subjects and yielded a pixel classification accuracy of 96.46%; the MAE of the infarction size was 22.31 mm^2. All of these results demonstrate that MuTGAN can aid in the clinical diagnosis of MI assessments.

References

1. Bijnens, B., Claus, P., Weidemann, F.: Investigating cardiac function using motion and deformation analysis in the setting of coronary artery disease. Circulation 116(21), 2453–2464 (2007).
2. Ordovas, K. G., Higgins, C. B.: Delayed contrast enhancement on MR images of myocardium: past, present, future. Radiology 261(2), 358–374 (2011).
3. Fox, C. S., Muntner, P.: Use of Evidence-Based Therapies in Short-Term Outcomes of ST-Segment Elevation Myocardial Infarction and Non–ST-Segment Elevation Myocardial Infarction in Patients With Chronic Kidney Disease. Circulation 121(3), 357–365 (2010).
4. Lipton, M. J., Bogaert, J., Boxt, L. M., Reba, R. C.: Imaging of ischemic heart disease. European radiology 12(5), 1061 (2001).
5. Wollmann, T., Ivanova, J., Gunkel, M.: Multi-channel Deep Transfer Learning for Nuclei Segmentation in Glioblastoma Cell Tissue Images. In: Bildverarbeitung für die Medizin 2018, pp. 316–321 (2018).
6. Xu, C., Zhang, H.: Direct detection of pixel-level myocardial infarction areas via a deep-learning algorithm. In: MICCAI. Springer, pp. 240–249 (2017).
7. Popescu, I. A., Irving, B., Borlotti, A.: Myocardial Scar Quantification Using SLIC Supervoxels-Parcellation Based on Tissue Characteristic Strains. In: STACOM. Springer, pp. 182–190 (2016).
8. Xue, W., Li, S.: Full quantification of left ventricle via deep multitask learning network respecting intra-and inter-task relatedness. In: MICCAI. Springer, pp. 276–284 (2017).
9. Ogawa, R., Kido, T.: Diagnostic capability of feature-tracking cardiovascular magnetic resonance to detect infarcted segments: a comparison with tagged magnetic resonance and wall thickening analysis. Clinical radiology 72(10), 828–834 (2017).
10. Xingjian, S. H. I., Chen, Z.: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: NIPS, 802–810 (2015).
11. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: CVPR, 3431–3440 (2015).
12. Graves, A., Schmidhuber, J.: Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks 18(5-6), 602–610 (2005).
13. Bleton, H., Margeta, J.: Myocardial infarct localization using neighbourhood approximation forests. In: STACOM. Springer, pp. 108–116 (2015).
14. Karim, R., Housden, R. J., Balasubramaniam, M.: Evaluation of current algorithms for segmentation of scar tissue from late gadolinium enhancement cardiovascular magnetic resonance of the left atrium: an open-access grand challenge. JCMR 15(1), 105 (2013).