# Left-Ventricle Quantification Using Residual U-Net

Eric Kerfoot, James Clough, Ilkay Oksuz(✉), Jack Lee, Andrew P. King, and Julia A. Schnabel

School of Biomedical Engineering & Imaging Sciences,
King's College London, London, UK
{eric.kerfoot,ilkay.oksuz}@kcl.ac.uk

**Abstract.** Estimating dimensional measurements of the left ventricle provides diagnostic values which can be used to assess cardiac health and identify certain pathologies. In this paper we describe our methodology of calculating measurements from left ventricle segmentations automatically generated using deep learning. We use a U-net convolutional neural network architecture built from residual units to segment the left ventricle and then process these segmentations to estimate the area of the cavity and myocardium, the dimensions of the cavity, and the thickness of the myocardium. Determining if an image is part of the diastolic or systolic portion of the cardiac cycle is done by analysing the cavity volume. The quality of our results are dependent on our training regime where we have generated a large derivative dataset by augmenting the original images with free-form deformations. Our expanded training set, in conjunction with simple affine image transforms, creates a sufficiently large training population to prevent over-fitting of the network while still creating an accurate and robust segmentation network. Assessing our method on the STACOM18 LVQuan challenge dataset we find that it significantly outperforms the previously published state-of-the-art on a 5-fold validation all tasks considered.

**Keywords:** Cardiac MR · Cardiac quantification · Convolutional neural networks

## 1 Introduction

Dimensional measurements of the left ventricle (LV) can be used to quantify cardiac health and identify some pathologies. For example, measuring the cavity volume throughout the cardiac cycle allows the calculation of the ejection fraction and stroke volume biomarkers, and measuring the thickness of the myocardium aids in diagnosing hypertrophic cardiomyopathy. Calculating these measurements automatically and accurately is an open challenge which, if addressed, would provide a time and cost saving tool for clinical diagnostic use.

In this paper we describe our method for left ventricle quantification based on the automatic segmentation of the LV using deep learning. We trained a

M. Pop et al. (Eds.): STACOM 2018, LNCS 11395, pp. 371–380, 2019.
https://doi.org/10.1007/978-3-030-12029-0_40

network to segment the myocardium of full cycle cardiac MR images (cMRI) captured at three positions on the LV. These segmentations are then processed to calculate, for each image, cavity and myocardium volumes, cavity dimensions, regional wall thicknesses, and assignment to diastolic or systolic portion of the cycle. Our method relies on an accurate and robust segmentation network and so we use data augmentation to create an expanded input dataset from an original smaller set, and apply further augmentations during the training process.

This work is our entry for the STACOM Left Ventricle Full Quantification Challenge for MICCAI 2018. The challenge involves the calculation of cavity and myocardial areas, three dimensional values of the cavity, six regional wall thickness values, and assigning each image to diastole or systole phase. We will discuss in this paper our method for inferring a myocardial segmentation from the training input images, and then the calculation of these metrics from the segmentations using image processing routines.

## 2 Background

The quantification of cardiac indices from cardiac MRI has been investigated with a variety of approaches in literature [7]. Early works relied on manual delineation of the myocardial borders and used these manual contours to calculate the cardiac indices [10]. Recent methods have bypassed the segmentation step and posed the problem as a direct regression of cardiac indices from cMRI images. The early efforts at regression focused on a two step strategy, where hand-crafted features are generated and used in a regression setting for cardiac indices estimation [14]. With the development of deep learning techniques more advanced models have been proposed to circumvent the problems of hand-crafted features by using end-to-end trained networks [16]. More recently, Xue et al. [17] have incorporated the temporal information from different phases of the cardiac cycle using a recurrent neural network.

Machine learning techniques to automate myocardial segmentation have achieved considerable success [7]. The U-net approach of Ronneberger et al. [8] has been widely adopted in the context of medical image analysis. In the context of cardiac MR segmentation, Avendi et al. [1] combined CNNs and a deformable model to perform LV segmentation, though only for the endocardial wall and only in the end-diastole and end-systole phases.

Tran [12] used a 15-layer CNN to perform complete left and right ventricular segmentation. Tan et al. [11] proposed to use radial distances calculated with a polar transform rather than per-pixel binary image segmentations. Most recently, Oktay et al. [6] incorporate global shape information into neural networks, and Bai et al. [2] showcases human-level performance for segmentation using a fully convolutional neural network.

# 3 Methods

The proposed framework consists of three stages: (1) image preprocessing and augmentation, (2) image segmentation with our CNN architecture, and (3) estimation the cardiac indices using the output segmentation information.

## 3.1 Image Preprocessing

An accurate and robust segmentation network requires a large dataset so that it learns a general solution which can correctly segment images not seen in training, and does not become over-fitted to the input data. The key concept is variety since the network is being trained to identify geometry embedded in varying contexts. If some ancillary feature in this context is often present with important features the network will correlate these features and produce a poor result for inputs lacking the ancillary feature. To provide this varied data, we first create an expanded input dataset from the original challenge data and then apply data augmentation during the training process.

The training dataset for the STACOM challenge consists of 145 subjects each with 20 image and segmentation frames, resulting in an input image matrix of dimensions 2900 × 80 × 80 pixels and a binary segmentation image of the same size. As a preprocessing step we apply a non-rigid deformation to each image/segmentation pair twenty times to produce an expanded image matrix and a segmentation matrix of dimensions 2900 × 21 × 80 × 80. This processing step deforms the pixels of the image/segmentation pair near the centre in randomised directions using a smooth interpolation function, thus producing an image which has slightly different geometry to its original but still has an accompanying segmentation which segments the myocardium.

During training, we use data augmentation [5,9] when creating an input batch of images at each step. Images from the expanded dataset are selected at random, then a random selection of transpose, flip, 90/180/270-degree rotation, and shift operations are applied to each image/segmentation pair. Applying these transformations essentially produces a further expanded dataset which contains increased image variation although not further geometric variation. This prevents the network from fixating on features in specific regions of its perceptive field since the random transformations move such features around the field during training. Augmentation is the only technique we use to prevent over-fitting, other techniques like dropout were found not to improve performance and so omitting them contributed to a simpler network architecture.

## 3.2 Segmentation Network Architecture

Our network architecture (Fig. 1) is based on U-net [8] with the layers of the encoding and decoding stages defined using residual units [4,18]. Each box of the encoding portion is implemented with a residual unit and labelled with the output volume shape. The decoding portion boxes are implemented with an upsample unit and have the same output volume dimensions as the encoding layer on the same level.

Parametric rectifying linear units [3] are used to allow the network to learn a better activation which improves segmentation. Instance normalization is used to prevent contrast shifting, ensuring input image contrast is not skewed by being batched with images having significantly different contrast ranges [13].

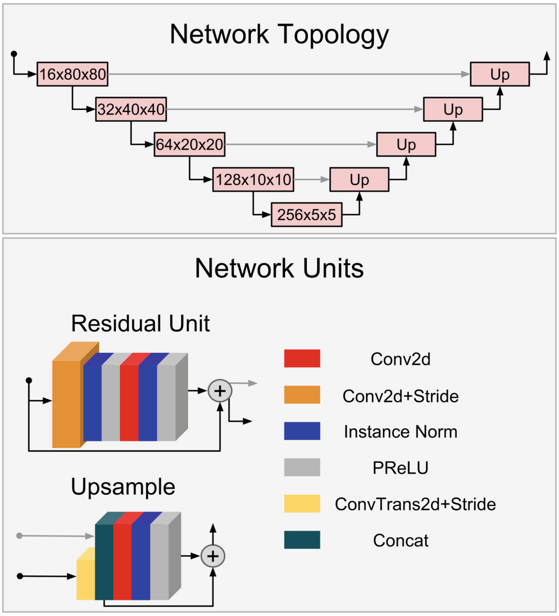

**Fig. 1.** U-net network architecture

As Fig. 1 shows, we use convolutions with a stride of 2 for downsampling and transpose convolutions with a stride of 2 for upsampling, rather than pooling and unpooling layers. This allows the network to learn optimal downsample/upsample operations while also reducing the number of layers in the network's units. Note that the first residual unit, marked 16 × 80 × 80, uses a stride of 1 in its first convolution layer so that the original input image is not immediately downsampled.

### 3.3 Cardiac Indices Estimation

All indices are calculated from myocardial segmentations derived from the CNN using conventional image processing routines. The myocardial area is estimated by counting the number of pixels in the segmentation image and scaling by the image dimensions. Similarly, the cavity area is estimated by counting pixels in the enclosed area of the segmentation image. Phase estimation is derived from the cavity areas, where the image having the smallest area is considered the end systole image, and the largest is the end diastole image.

Calculating the three cavity dimension values (IS-AL, I-A, and IL-AS) and the six regional wall thickness values (IS, I, IL, AL, A, and AS) requires measuring thickness values over an area of the segmentation image. Our technique is to unravel the segmentation by converting each pixel coordinate into a polar coordinate then projecting those coordinates onto a Cartesian plane. Figure 2 illustrates this process by unravelling a shaded annulus into an image where, for each pixel, the row index represents the distance from the original image centre and the column index represents the angle in degrees around the centre clockwise starting from the top.

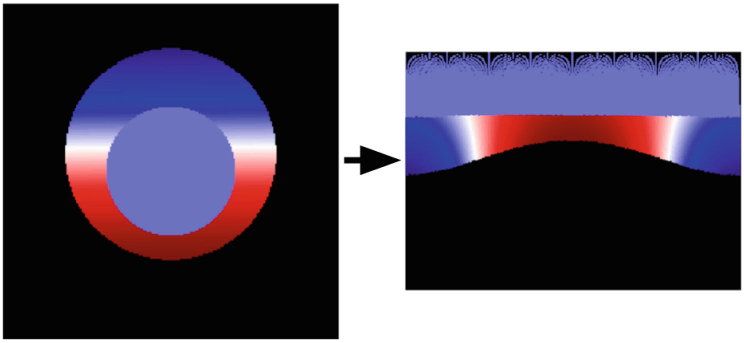

**Fig. 2.** Segmentation polar projection showing transformation of original image to unravelled equivalent image (colouration chosen to highlight transformation). (Color figure online)

By unravelling the segmentation images in this way, the cavity dimensions can be calculated by averaging the maximal row index of non-zero pixels over each column corresponding to the angular areas of the segmentation to be considered. For example, to calculate the IS-AL cavity dimension, the average of the last row index of non-zero pixels is calculated over columns 30 to 90 and 210 to 270 then multiplied by two. Similarly the region wall thicknesses are calculated by averaging the difference between the last row index of non-zero pixels and the first over the columns to consider for each measurement zone.

### 3.4 Implementation Details

We trained our network using the PyTorch 0.4.0 library on nVidia TITAN Xp and P6000 GPUs. Segmentation analysis routines are all implemented in Python 3.6 using the Numpy, SciPy, and Numba libraries. The training takes approximately 3 h for 10000 steps with a batch size of 1200. Our code is available to download at https://github.com/ericspod/STACOM18.

# 4 Experimental Results

To demonstrate the robustness of our network and analysis routines, we trained the network on three different dataset folds where the set of subjects is divided into three, five, and seven groups. We separately trained fifteen networks on each dataset instance with one of these groups reserved as the validation set, each network trained for 10000 steps with a batch size of 1200.

This batch size was chosen as the largest we could fit on one of our GPUs. The folds were created using the recommended process of the STACOM LVQuan18 challenge, ensuring that data from one subject is not split between training and test sets, and facilitating comparison with other methods in the challenge.

## 4.1 Qualitative Results

In Fig. 3, we show an example image from the training dataset with the corresponding segmentation mask generated using our algorithm. The segmentation mask is very accurate and has only few over- or under-segmented regions. The successful segmentations generated by our algorithm is the key for cardiac indices estimation.

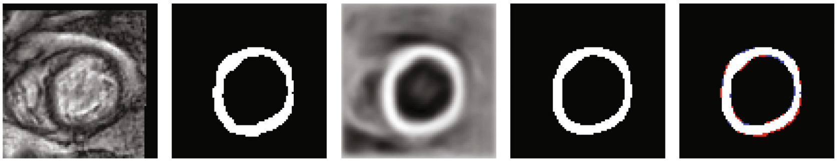

**Fig. 3.** Example training images: first and second images are input image and associated ground truth segmentation from training dataset, third is output logits from network, fourth is estimated segmentation from thresholding logits, fifth illustrates difference between ground truth and estimated segmentation, where red represents the under segmented regions and blue the over-segmented regions. (Color figure online)

## 4.2 Quantitative Results

Table 1 lists the results of applying the trained networks to the validation set and then applying our metric routines to calculate the area, dimension, wall thickness, and phase values for each image. The table lists, for each metric type, the mean absolute error (MAE) and Pearson correlation coefficient ($\rho$) when comparing the ground truth metrics against the calculated values. The last column lists the intersection-over-union error between the ground truth myocardial segmentation images versus the calculated segmentations. The last row includes the results of applying our methodology to the challenge's test dataset, omitting IoU as the ground truth segmentations are not yet available.

The conclusion from this study is that our network is generally stable and insensitive to varying training input datasets, exhibiting a loss of accuracy correlated with fold size. In particular the IoU values for fold three are greater than

**Table 1.** Fold validation results. The area, cavity dimension, and region wall thickness metrics were calculated for the validation portion of the fold (e.g. for fold 3,0 the first third of images from the dataset were validation). The mean absolute error (MAE, in mm$^2$ units for area and mm units for dim/RWT) and Pearson correlation coefficient ($\rho$) were calculated for each metric as a group. The IoU column is the intersection-over-union difference between the segmentations inferred from the fold network and the ground truth segmentations.

| Fold | | Area | | Cavity Dim. | | RWT | | Phase | IoU |
|---|---|---|---|---|---|---|---|---|---|
| | | MAE | $\rho$ | MAE | $\rho$ | MAE | $\rho$ | % Error | |
| 3 | 0 | 93.184 | 0.9631 | 1.0112 | 0.9808 | 0.8864 | 0.9355 | 6.758 | 0.1218 |
| | 1 | 71.872 | 0.9827 | 0.8712 | 0.9864 | 0.7352 | 0.9565 | 7.379 | 0.1214 |
| | 2 | 79.744 | 0.9659 | 0.8784 | 0.9836 | 0.7880 | 0.9416 | 6.931 | 0.1274 |
| Mean | | 81.600 | 0.9706 | 0.9203 | 0.9836 | 0.8032 | 0.9445 | 7.023 | 0.1235 |
| 5 | 0 | 64.576 | 0.9841 | 0.7840 | 0.9875 | 0.7008 | 0.9597 | 6.482 | 0.1059 |
| | 1 | 58.368 | 0.9873 | 0.7808 | 0.9883 | 0.6800 | 0.9621 | 6.689 | 0.1078 |
| | 2 | 56.512 | 0.9879 | 0.7728 | 0.9885 | 0.6456 | 0.9641 | 6.965 | 0.1101 |
| | 3 | 68.928 | 0.9793 | 0.8488 | 0.9861 | 0.7056 | 0.9551 | 7.172 | 0.1125 |
| | 4 | 63.104 | 0.9837 | 0.7712 | 0.9903 | 0.6776 | 0.9614 | 6.310 | 0.1100 |
| Mean | | 62.298 | 0.9845 | 0.7915 | 0.9881 | 0.6819 | 0.9605 | 6.724 | 0.1093 |
| 7 | 0 | 59.648 | 0.9863 | 0.7912 | 0.9878 | 0.6640 | 0.9635 | 6.344 | 0.1031 |
| | 1 | 55.488 | 0.9887 | 0.7496 | 0.9896 | 0.6352 | 0.9662 | 6.482 | 0.1053 |
| | 2 | 64.768 | 0.9870 | 0.8336 | 0.9900 | 0.6560 | 0.9703 | 6.758 | 0.1038 |
| | 3 | 64.704 | 0.9879 | 0.7720 | 0.9900 | 0.6456 | 0.9694 | 6.137 | 0.1057 |
| | 4 | 54.080 | 0.9867 | 0.7416 | 0.9871 | 0.6328 | 0.9615 | 7.379 | 0.1032 |
| | 5 | 54.912 | 0.9879 | 0.7096 | 0.9905 | 0.6112 | 0.9689 | 6.724 | 0.1015 |
| | 6 | 62.528 | 0.9886 | 0.7360 | 0.9913 | 0.6576 | 0.9678 | 6.517 | 0.1048 |
| Mean | | 59.447 | 0.9875 | 0.7619 | 0.9895 | 0.6432 | 0.9668 | 6.620 | 0.1039 |
| Test | | 301.811 | 0.9384 | 2.3252 | 0.9742 | 2.1531 | 0.7789 | 10.0 | N/A |

those for fold seven as one would expect given that the former had fewer training images and more unseen validation images. However the IoU values within folds are roughly equivalent, demonstrating the training was not sensitive to the presence of absence of particular input images.

The results from the test dataset correlate in general with this conclusion although there is a meaningful loss of accuracy across all metrics. For a challenge of this type it is expected that the test set would be particularly difficult and the results show this, in particular we observed some images of the test set could not be segmented correctly and so our algorithms had to compensate for an incomplete annular segmentation to generate a result.

## 5 Discussion and Conclusion

In this paper we have proposed our CNN-based approach to calculating left-ventricle metrics for the STACOM 2018 challenge. The central component to this approach is a U-net based network for segmenting the myocardium of the left ventricle. Our modifications to the classic U-net architecture and our training process are critical components to producing an accurate segmentation. Segmentations generated by this network are then processed to calculate the metrics in question.

We chose a method which relies on segmentation, in contrast to [15] which proposes a regression approach for calculating metrics directly with a RNN, for a number of practical and quantitative reasons. U-net and other architectures as applied to cardiac segmentation in particular are well understood and known to be robust when trained with a varied dataset. In particular residual units and data augmentation are now common practice when training such networks and have obvious benefits in terms of training speed and producing a network which is robust against input dissimilar from the training set.

More quantitatively, using segmentations allows the numeric and visual assessment of the network's outputs. Since the segmentations are used as input to routines for calculating the metrics, which are accurate when applied to the ground truth segmentations, the accuracy of the final metrics relies entirely on that of the segmentations. To determine if the network is producing accurate segmentations, metrics such as intersection-over-union can be used to assess accuracy, or image processing routines can be used to ensure each output image represents a single correct annulus. Most easily a simple visual inspection can be made of the outputs as overlaid on the source images.

In our experience it was clear when subjects were segmented poorly by inspecting the produced myocardial segmentations and it was equally clear why input image quality caused the network to fail. If images dissimilar from the training data produce poor results when applied to a network then visual analysis can help in diagnosing why, for example determining if other image structures were identified as the LV, if image noise made the myocardium indistinct, or if the network suffers from overfitting and so produced indistinct and jumbled segmentations. This contributes to a methodology which is less of a black-box approach compared to others, and which permits visual inspection and analysis when failure occurs.

Our scheme of image augmentations was designed to prevent overfitting to the set of training images, and so make our method more generalisable to the images in the test set (for which we do not have ground-truth segmentations due to the nature of the STACOM 2018 challenge). We assessed the effect of these augmentations by training the network with just the original training data and applying it to the test set images to produce segmentations.

Although we do not have ground-truth segmentations to compare with, visual inspection of the results, in Fig. 4 demonstrates the usefulness of the augmentation scheme. Furthermore, we counted the number of non-annular segmentations as a rough measure of segmentation quality. Overall, the test set yielded

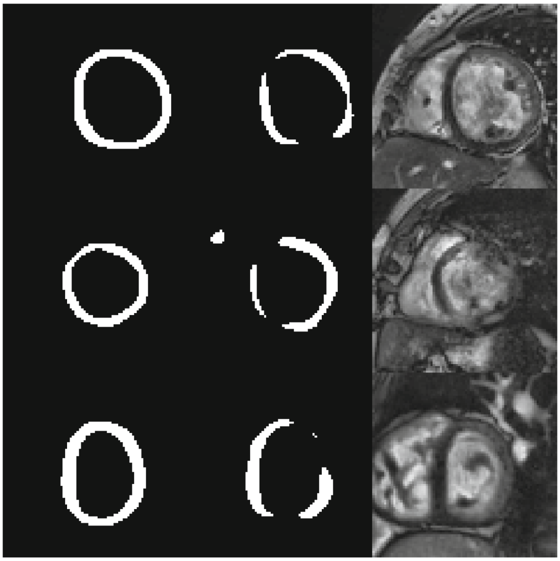

**Fig. 4.** Example of bad segmentations. Mask on left is produced by network trained with augments, mask on right without. Right-hand side is source image from test dataset.

6 non-annular segmentations when the network was trained with our augmentations, and 158 such segmentations when trained without the augmentations, demonstrating that they are key to training a network which generalises well.

Corollary to this is the observation that training without augmentations produces a much smaller final loss value as compared to training with augmentations. Having trained on the whole dataset for 5000 iterations with a batch size of 600 images, the final loss value without augmentations is 0.0092 but with augmentations is 0.0575. In conjunction with the observation of how manybreak non-annular segmentations are produced, this clearly indicates significant overfitting to the training dataset when augmentations are not used.

**Acknowledgments.** This work was supported by an EPSRC programme Grant (EP/P001009/1) and the Wellcome EPSRC Centre for Medical Engineering at the School of Biomedical Engineering and Imaging Sciences, King's College London (WT 203148/Z/16/Z). The GPUs used in this research was generously donated by the NVIDIA Corporation.

## References

1. Avendi, M., Kheradvar, A., Jafarkhani, H.: A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. Med. Image Anal. **30**, 108–119 (2016)
2. Bai, W., et al.: Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. J. Cardiovasc. Magn. Reson. **20**, 65 (2018)

3. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on imageNet classification. CoRR abs/arXiv:1502.0185 (2015)
4. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. CoRR abs/1603.05027 (2016)
5. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems, NIPS 2012, vol. 1, pp. 1097–1105. Curran Associates Inc., USA (2012)
6. Oktay, O., et al.: Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. IEEE Trans. Med. Imaging **37**(2), 384–395 (2018)
7. Peng, P., Lekadir, K., Gooya, A., Shao, L., Petersen, S.E., Frangi, A.F.: A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. Magn. Reson. Mater. Phys. Biol. Med. **29**(2), 155–195 (2016)
8. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. CoRR abs/1505.04597 (2015)
9. Simard, P.Y., Steinkraus, D., Platt, J.: Best practices for convolutional neural networks applied to visual document analysis. Institute of Electrical and Electronics Engineers, Inc. August 2003
10. Suinesiaputra, A., et al.: Quantification of LV function and mass by cardiovascular magnetic resonance: multi-center variability and consensus contours. J. Cardiovasc. Magn. Reson. **17**(1), 63 (2015)
11. Tan, L.K., Liew, Y.M., Lim, E., McLaughlin, R.A.: Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine MR sequences. Med. Image Anal. **39**, 78–86 (2017)
12. Tran, P.V.: A fully convolutional neural network for cardiac segmentation in short-axis MRI. arXiv preprint arXiv:1604.00494 (2016)
13. Ulyanov, D., Vedaldi, A., Lempitsky, V.S.: Instance normalization: The missing ingredient for fast stylization. CoRR abs/1607.08022 (2016). http://arxiv.org/abs/1607.08022
14. Wang, Z., Salah, M.B., Gu, B., Islam, A., Goela, A., Li, S.: Direct estimation of cardiac biventricular volumes with an adapted bayesian formulation. IEEE Trans. Biomed. Eng. **61**(4), 1251–1260 (2014)
15. Xue, W., Brahm, G., Pandey, S., Leung, S., Li, S.: Full left ventricle quantification via deep multitask relationships learning. Med. Image Anal. **43**, 54–65 (2018)
16. Xue, W., Islam, A., Bhaduri, M., Li, S.: Direct multitype cardiac indices estimation via joint representation and regression learning. IEEE Trans. Med. Imaging **36**(10), 2057–2067 (2017)
17. Xue, W., Nachum, I.B., Pandey, S., Warrington, J., Leung, S., Li, S.: Direct estimation of regional wall thicknesses via residual recurrent neural network. In: Niethammer, M., et al. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 505–516. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_40
18. Zhang, Z., Liu, Q., Wang, Y.: Road extraction by deep residual U-Net. CoRR abs/1711.10684 (2017)