Ejection Fraction Estimation Using a Wide Convolutional Neural Network

AbdulWahab Kabani and Mahmoud R. El-Sakka(✉)

Department of Computer Science, The University of Western Ontario,
London, ON, Canada
{akabani5,melsakka}@uwo.ca

Abstract. We introduce a method that can be used to estimate the ejection fraction and volume of the left ventricle. The method relies on a deep and wide convolutional neural network to localize the left ventricle from MRI images. Then, the systole and diastole images can be determined based on the size of the localized left ventricle. Next, the network is used in order to segment the region of interest from the diastole and systole images. The end systolic and diastolic volumes are computed and then used in order to compute the ejection fraction. By using a localization network before segmentation, we are able to achieve results that are on par with the state-of-the-art and by annotating only 25 training subjects (5% of the available training subjects).

Keywords: Localization · Detection · Recognition · Artificial neural networks · Deep learning · Convolutional neural network · Image classification · Cardiac MRI · Left ventricle · Automatic ejection fraction estimation

1 Introduction

The ejection fraction is a measure of the outbound blood pumped from the heart after each heartbeat. The ejection fraction, and the systolic and diastolic volumes are important measures that can help cardiologists assess how healthy the heart is. Manually estimating the end-systolic volume, the end-diastolic volume, and the ejection fraction from cardiac MRI images is a time consuming process. A cardiologist or an expert has to manually segment all slices before the volume can be calculated. We propose a deep and wide convolutional neural network that can be used to localize the left ventricle from an MRI image. This network strikes a nice balance between performance and hardware requirement. Then, the systole and diastole images can be determined based on the size of the localized left ventricle. Next, the same network can be used to segment the cavity in the left ventricle. Using the DICOM meta fields, we can compute the volume size.

Left ventricle volume estimation has been an active area of research [4,9,14,15]. Recently CNNs were used in order to tackle this problem [7,13,27]. For instance, in [27] a U-Net [19] network was used in order to segment the left ventricle.

F. Karray et al. (Eds.): ICIAR 2017, LNCS 10317, pp. 87–96, 2017.
DOI: 10.1007/978-3-319-59876-5_11

A Convolutional Neural Network (CNN) is a neural network that contains some layers with restricted connectivity. CNNs were introduced in [3] and achieved good results on the MNIST data set [12]. The success of CNNs on the MNIST [12] was reproduced on image classification problems. CNNs can now produce state of the art performance in many classification tasks and on challenging datasets such as [1,20]. The success of CNNs is due to many reasons including large training data sets [1,20], powerful hardware, regularization techniques such as Dropout [6,24], initialization methods [5], ReLU activations [17], and data augmentation. Since 2012, many networks that can perform image recognition were introduced [11,22,25]. Recently, CNNs have been used to perform detection and localization [2,19,21,23,26].

CNNs typically require large training data sets in order to be successful. However, cleaning and standardizing the data can help a lot in alleviating this requirement. We use a dataset of MRI images provided by the National Heart, Lung, and Blood Institute (NHLBI) [8]. The dataset [8] contains 500 training studies, 200 validation studies, and 440 testing studies. This dataset includes the MRI images along with the diastole and systole volumes. The dataset does not contain the left ventricle segmentation. Therefore, the left ventricle in the training data should be manually segmented and annotated to be able to train a CNN that can segment left ventricle. Manually annotating the training data is a time consuming process. Therefore, we first train a model to localize the left ventricle which is a much easier task. To perform this step, one image from each subject is annotated and used to train a localization network. Once trained, this localization network is used to localize the left ventricle in all the images. Once the left ventricle is localized, the task of training a model to segment the left ventricle becomes much easier and requires a very small amount of annotated training data. For the segmentation, we annotated only 25 training studies out of 500 training studies (5% of the number of studies). Finally, after segmenting the left ventricle, the systolic and diastolic volumes can be easily calculated. The results are reported on the whole testing set, which includes 440 subjects. The test set was never manually annotated in any way.

In this paper, we divide the task into several steps in order to ensure that the training process is successful. First, the left ventricle is localized (Sect. 2). Then, the systole and diastole images for each slice are determined. Next, the left ventricle segmentation process is described in Sect. 3. Finally, the volume of the left ventricle is estimated (Sect. 4). In Sect. 5, we present the results. We conclude our work in Sect. 6. An overview of our proposed solution is shown in Fig. 1.

2 Localization

The dataset contains (500 training subjects, 200 validation subjects, and 440 testing subjects). Each subject includes 8–16 short axis views (or slices). These slices provide a view of the heart at different levels. Each slice contains a set of (30 images or less) spanning one cardiac cycle. We are interested in only

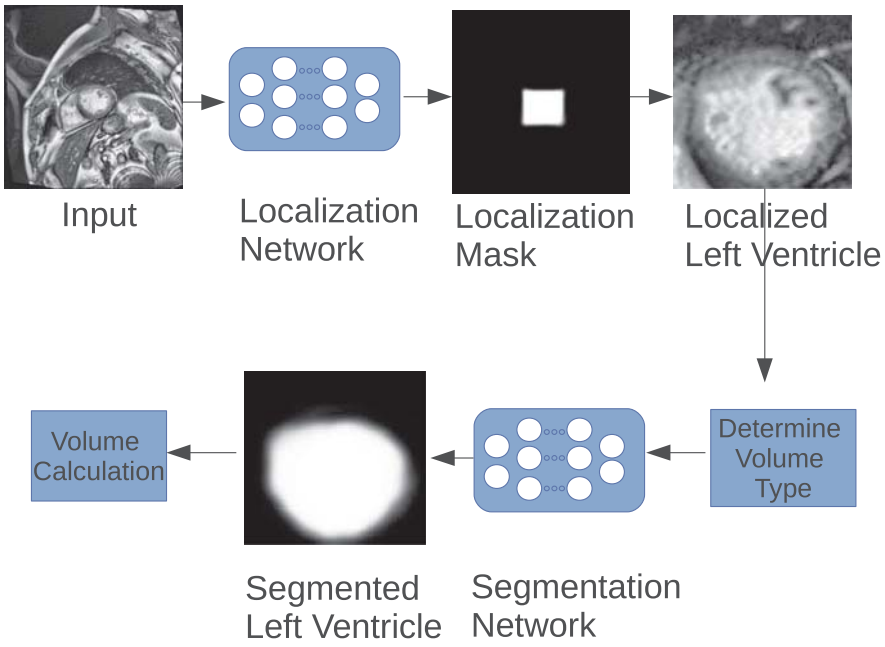

Fig. 1. Overview: a set of training images (1 per subject) are used to train the localization network. Then, the network is used to predict the localization mask. The localization mask is thresholded using Otsu [18]. Once the left ventricle is localized, we can determine the systole and diastole images. The left ventricle is segmented by training a network on only 25 training subjects. After thresholding, the area of the left ventricle corresponds to the sum of pixel values that are not 0. The volume can be calculated by multiplying each area by the slice height and summing up the volumes.

two images in each slice (the end-systolic and end-diastolic images). In each slice, these two images show us when the left ventricle is fully contracted and expanded, respectively. In order to determine these two phases, we first need to localize the left ventricle.

We trained a localization network on a set of randomly sampled images from the training set. We made sure that one image is sampled from each patient. The architecture of the localization network is described in Fig. 2. As shown in Fig. 2, the input image is re-sized to (128, 128). Then, it is passed to the network and it goes through a set of convolutional layers. Next, the last layers from each block -except for the first one- are upsampled and merged. Three types of merging is done: type_1 involves merging (blocks_2, blocks_3, blocks_4, blocks_5), type_2 involves merging (blocks_3, blocks_4, blocks_5), and finally type_3 merging (block_4, and block_5). It is worth noting that type_1 carries more high resolution information while type_3 carries highly abstract information but with less resolution (due to downsampling).

Each convolutional layer is followed by rectified linear units (ReLU) activation [17]. The output layer has a sigmoid activation to ensure that each pixel in the output mask ranges between 0 (black) to 1 (white). Maxpooling is used in order to extract more abstract features. We opted for upsampling by repeating

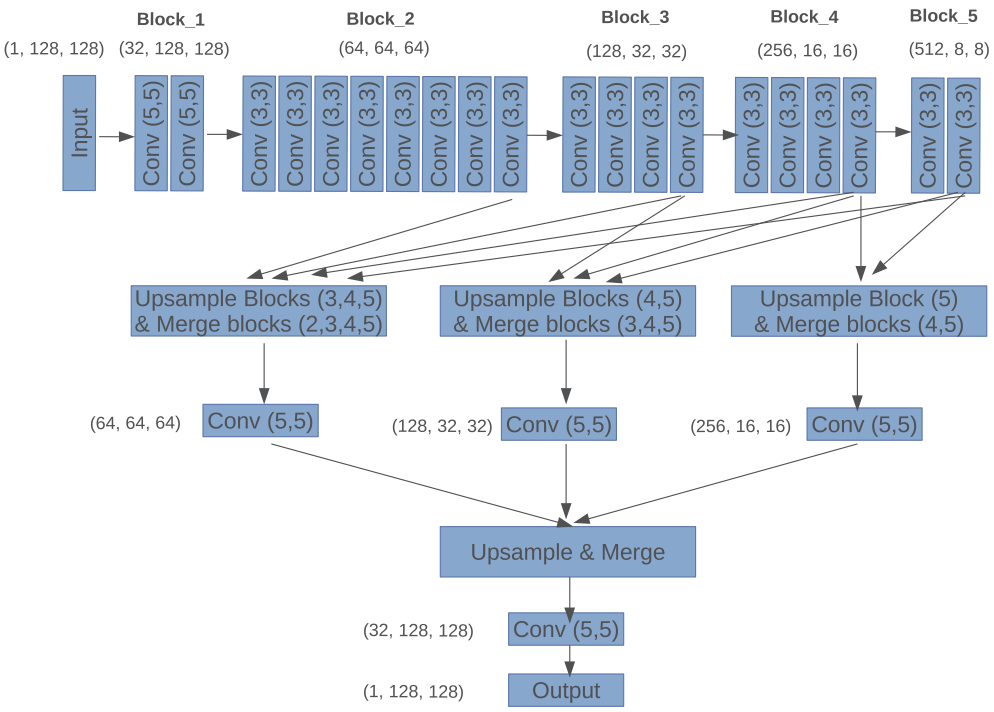

Fig. 2. Network architecture: this figure shows the architecture of one of the networks that were used for localization and segmentation of the left ventricle. The input goes through a set of convolutional layers. Then, the last layer from each block (except for block_1) are upsampled and merged in a hierarchical fashion in order to get the output mask in the original resolution.

the units rather than through parametrized upsampling in order to reduce the training time and GPU RAM consumption. This network strikes a nice balance between performance and hardware requirement. Indeed, this network was also tested on a laptop with around 3.5 GB of available GPU RAM.

The only input pre-processing that we did was subtracting the mean input of each channel and then dividing by the standard deviation. It is worth mentioning that in order for the left ventricle area to be determined correctly, we resized all images using the pixel spacing DICOM field. This ensures that each pixel represents 1 mm. However, when training the localization network, we resized the input image to (128, 128) because the network can only accept fixed sizes. Once the network is used to predict the mask, the mask is resized to the original size where 1 $pixel = 1$ mm. The following equations show how the pixel spacing is used to resize the image:

$$w_{new} = w_{old} \times \Delta s \tag{1}$$

$$h_{new} = h_{old} \times \Delta s \tag{2}$$

where w_{old} is the old image width, w_{new} is the new image width, h_{old} is the old image height, h_{new} is the new image height, and Δs is the pixel spacing, respectively.

The output of this step is shown in Fig. 3.

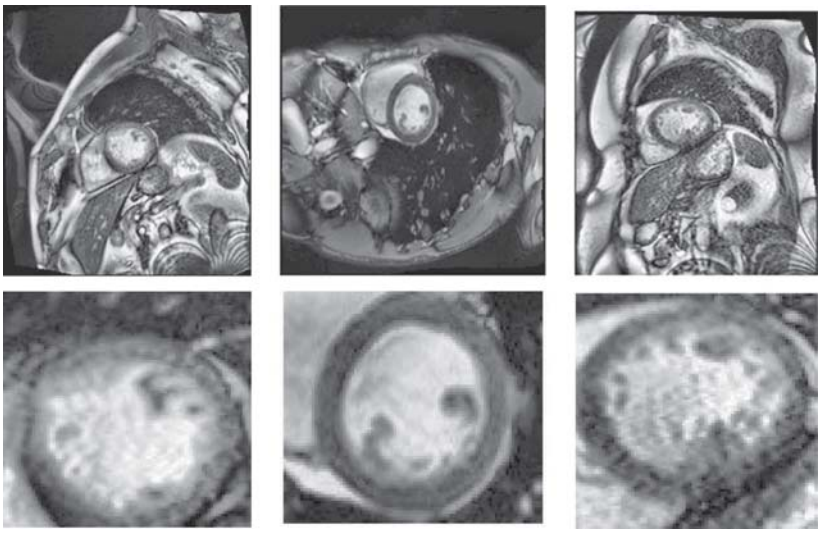

Fig. 3. Localization sample: first row shows the original images. Second row shows the localized left ventricle.

3 Determining Images of Interest and Segmentation

Because the size of the left ventricle varies between the end-systolic and end-diastolic phases, we can determine these phases by the size of the white area in the masks. For each slice, we consider the image with the smallest mask as end-systolic and the one with the largest mask as end-diastolic. A neural network with the same architecture as shown in Fig. 2 can be used to segment the left ventricle. The network is trained only on the end-sysolic and end-diastolic images because these are the images that are needed to compute the end-systolic and end-diastolic volumes.

It should be noted that while the network has a similar architecture to the localization network, the input size for the segmentation network is smaller (80×80) vs (128×128) for the localization network. This is because the input images are much smaller than the original images. Consequently, the segmentation network is trained much faster than the localization network. The segmentation network was trained only on 540 images representing 25 subjects. The training set contains 500 training subjects. We could not use the whole training set because the process of performing manual segmentation annotation is very time consuming. Because of the localization process, training the segmentation network on such a small data is possible as most of the irrelevant pixels are removed before segmentation.

Figure 4 shows a sample of left ventricle images and the predicted segmentation masks. These masks were generated by training an ensemble of 10 segmentation networks. The ensemble contains networks similar to the one described in Fig. 2 but with varying the number of layers and kernel size in each block. We found that training an ensemble helps a lot since the number of training images is very low.

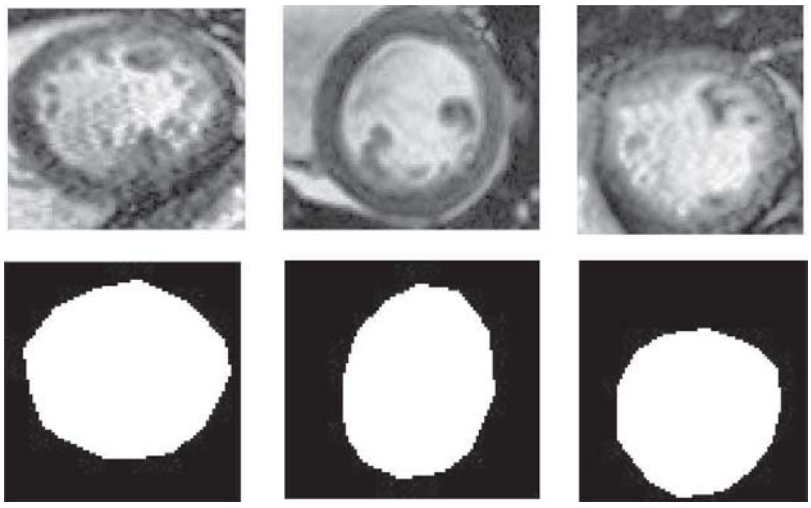

Fig. 4. Segmentation sample: first row shows the left ventricle images. Second row shows the segmentation mask.

4 Volume Estimation

At this stage, we can calculate the volume because the left ventricle is segmented. We calculated the volume using the following equation:

$$V = \sum_{i=0}^{N} SA_i \times SH_i \tag{3}$$

where SA_i and SH_i are the area and height of the slice i, respectively. The area of the slice is calculated by taking the sum of all active pixels (nonzero pixels) in the segmentation mask. SH_i is the result of taking the absolute difference of slice locations from two consecutive slices.

Equation 4 shows how to compute the ejection fraction which is a measure of the outbound blood pumped from the heart after each heartbeat.

$$EF = 100 \times \frac{(V_D - V_S)}{V_D} \tag{4}$$

where V_D and V_S are the end diastolic and end systolic volumes, respectively.

Because we did not train on the whole training set, we train four linear regression and random forest models to improve the results of the volume calculation. The features that are used to train these models include the systole and diastole volumes along with other features such as average slice height, average slice area, etc. Next section, we will report the results that we achieved.

5 Results

We trained the model on a computer with titan X (12 GB RAM). It is worth mentioning that the networks that we described in this paper were also tested on a laptop with a 3.5 GB of available GPU RAM.

Table 1. Diastole RMSE comparison with other state of the art solutions. The state of the art solutions were obtained from [16]

Method	Diastolic RMSE (ml)
Tencia Woshialex [13]	12.02
Ours	**13.37**
JuliandeWit [27]	13.63
Kunsthart [10]	13.65
ShowMeTheMoney	13.2

Table 2. Systole RMSE comparison with other state of the art solutions. The state of the art solutions were obtained from [16]

Method	Systolic RMSE (ml)
ShowMeTheMoney	9.31
Tencia Woshialex [13]	10.19
JuliandeWit [27]	10.32
Kunsthart [10]	10.43
Ours	**12.1**

Table 3. Ejection fraction RMSE comparison with other state of the art solutions. The state of the art solutions were obtained from [16]

Method	Ejection fraction RMSE (%)
ShowMeTheMoney	4.69
Tencia Woshialex [13]	4.88
JuliandeWit [27]	5.04
Ours	**5.97**
Kunsthart [10]	6.99

Training the localization network for 500 epochs with learning rate 0.001 can take around 4.1 h. On the other hand, training the segmentation network for 100 epochs takes only 28 min. Since we trained an ensemble of 10 segmentation networks, the segmentation training process takes around 5 h. The mean absolute value error (MAE) we were able to achieve for the end-diastolic, end-systolic, and ejection fraction are 9.94 ml, 8.42 ml, and 4.47%, respectively. On the other hand, the RMSE errors are slightly higher (13.37 ml for the diastolic volume, 12.1 for the systolic volume and 5.97% for the ejection fraction). According to [16], these values are comparable to the performance of humans in estimating the volumes. The differences between two humans in estimating the end diastolic, end systolic, and ejection fraction are: 13 ml, 14 ml, and 6%, respectively. Tables 1, 2, and 3 show a comparison between the results we achieved and the ones achieved

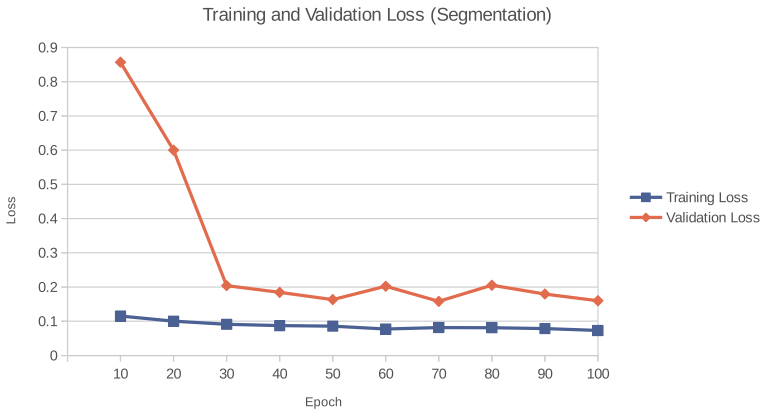

Fig. 5. Segmentation loss: the training and validation loss while training one of the segmentation models. There's a relatively big gap between the training and validation loss. This indicates that the model is likely overfitting the data.

by other methods. These results are very good considering the fact that the localization and segmentation CNNs were trained only on a small subset of the training set. Figure 5 shows the training and validation losses while training one of the segmentation models. Because of the gap between the two losses, it is very likely that the model is overfitting the data. This issue is likely because we only annotated 25 training subjects and used it to train the model. The results are likely to be improved if we annotate more subjects and train the model again.

6 Conclusion

We introduced a network that can be used to localize a region of interest in cardiac MRI images. Once the region of interest (left ventricle) is localized, the systole and diastole images are determined. Next, we segmented the left ventricle using the CNN we used for localization. Finally, we performed volume and ejection fraction estimation using the DICOM fields.

Acknowledgements. This research is partially funded by the Natural Sciences and Engineering Research Council of Canada (NSERC). This support is greatly appreciated. We would also like to thank kaggle, Booz Allen Hamilton, and the National Heart, Lung, and Blood Institute (NHLBI) for providing the MRI images.

References

1. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 248–255. IEEE (2009)
2. Erhan, D., Szegedy, C., Toshev, A., Anguelov, D.: Scalable object detection using deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2147–2154 (2014)

3. Fukushima, K.: Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern. **36**(4), 193–202 (1980)
4. Germano, G., Kiat, H., Kavanagh, P.B., Moriel, M., Mazzanti, M., Su, H.T., Train, K.F.V., Berman, D.S.: Automatic quantification of ejection fraction from gated myocardial perfusion spect. J. Nucl. Med. **36**(11), 2138 (1995)
5. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: International Conference on Artificial Intelligence and Statistics, pp. 249–256 (2010)
6. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580 (2012)
7. Kabani, A.W., El-Sakka, M.R.: Estimating ejection fraction and left ventricle volume using deep convolutional networks. In: Campilho, A., Karray, F. (eds.) ICIAR 2016. LNCS, vol. 9730, pp. 678–686. Springer, Cham (2016). doi:10.1007/978-3-319-41501-7_76
8. Kaggle: Data science bowl cardiac challenge data. https://www.kaggle.com/c/second-annual-data-science-bowl. Accessed 19 Mar 2016
9. Kaus, M.R., von Berg, J., Weese, J., Niessen, W., Pekar, V.: Automated segmentation of the left ventricle in cardiac MRI. Med. Image Anal. **8**(3), 245–254 (2004)
10. Korshunova, I.: Diagnosing heart diseases with deep neural networks. http://irakorshunova.github.io/2016/03/15/heart.html. Accessed 01 Feb 2017
11. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105 (2012)
12. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
13. Lee, T., Liu, Q.: Solution to win the second annual data science bowl. https://github.com/woshialex/diagnose-heart. Accessed 01 Feb 2017
14. Lin, X., Cowan, B.R., Young, A.A.: Automated detection of left ventricle in 4D MR images: experience from a large study. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 728–735. Springer, Heidelberg (2006). doi:10.1007/11866565_89
15. Lynch, M., Ghita, O., Whelan, P.F.: Automatic segmentation of the left ventricle cavity and myocardium in MRI data. Comput. Biol. Med. **36**(4), 389–407 (2006)
16. Mulholland, J.: Leading and winning team submissions analysis. http://www.datasciencebowl.com/leading-and-winning-team-submissions-analysis/. Accessed 04 Aug 2016
17. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML 2010), pp. 807–814 (2010)
18. Otsu, N.: A threshold selection method from gray-level histograms. Automatica **11**(285–296), 23–27 (1975)
19. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). doi:10.1007/978-3-319-24574-4_28
20. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. **115**(3), 211–252 (2015)

21. Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y.: OverFeat: integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229 (2013)
22. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
23. Song, H.O., Girshick, R., Jegelka, S., Mairal, J., Harchaoui, Z., Darrell, T.: On learning to localize objects with minimal supervision. arXiv preprint arXiv:1403.1024 (2014)
24. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
25. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. arXiv preprint arXiv:1409.4842 (2014)
26. Szegedy, C., Toshev, A., Erhan, D.: Deep neural networks for object detection. In: Advances in Neural Information Processing Systems, pp. 2553–2561 (2013)
27. de Wit, J.: Third place solution for the second kaggle national datascience bowl. https://github.com/juliandewit/kaggle_ndsb2. Accessed 01 Feb 2017