Assisting Radiologists in X-Ray Diagnostics

Cristian Avramescu, Bercean Bogdan, Stefan Iarca(✉), Andrei Tenescu, and Sebastian Fuicu

Politehnica University of Timisoara,
Piata Victoriei Nr. 2, 300006 Timisoara, Romania
stefan.iarca@student.upt.ro, sebastian.fuicu@cs.upt.ro

Abstract. Studies have shown that radiologists working together with Computer Aided Diagnostic software have increased accuracy. Automated screening software can be used to prioritize X-Rays coming in for diagnosis. We developed a suite of machine learning algorithms that aim to improve radiologist performance. It provides suggested diagnostics, a heatmap showing pathological areas and a bone subtracted version of the image which helps radiologists to identify fractures. We test different configurations for our diagnosis model, training it on both normal and enhanced images, using one or two branches. Our experiments show that adding enhanced inputs (lung segmented and bone subtracted versions of the input) increases the performance of our algorithm, which in turn increases the performance of the radiologist user. This shows that preprocessing the images before input increases model performance. More research is needed to find other preprocessing techniques, to refine existing ones, and to determine the optimal number and type of input X-Rays.

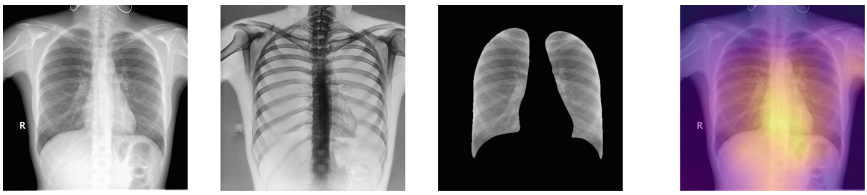
Keywords: Radiology · Deep Learning · X-ray · Segmentation · Bone · GAN

1 Introduction

In recent years, the decreasing costs and increasing efficiency of medical imaging equipment have made radiology a central tool in diagnosing diseases all around the world. X-rays are the most common form of medical imaging, with about 3.6 billion procedures performed yearly [1]. This large number can overcrowd radiologists, which decreases their diagnostic accuracy and can lead to fatigue [2].

We develop a suite of algorithms that aid radiologists in analyzing x-rays, by: screening the image and if any pathologies are found attempting to classify them into one of 13 classes, generating a heat-map of the diagnostic, lung segmenting the image, and creating a bone subtracted version of the image, which highlights the bone structure of the patient, making it easier to identify fractures.

N. M. Garcia et al. (Eds.): HealthyIoT 2019, LNICST 314, pp. 108–117, 2020.
https://doi.org/10.1007/978-3-030-42029-1_8



(a) Original x-ray (b) Bone Subtracted (c) Lung Segmented (d) Resulting CAM

Fig. 1. Preprocessing steps and resulting class activation map

1.1 Related Work

CheXNet, described in [3], was not the first published approach to classifying diseases in X-rays. Many different network architectures have been developed for this task, such as CNN+RNN network, for exploiting the dependencies between labels [4], Cascading CNN with different loss functions [5], Attention Guided CNN [6].

The algorithms mentioned above use the ChestX-ray14 dataset, made available by [7], the first attempt of offering a large, labeled x-ray public dataset. Two new improved datasets have been published since: CheXpert [8] and MIMIC-CXR [9], totaling around 590.000 frontal and lateral chest x-rays.

Automatic lung segmentation has been studied for a long time. [10] used a gray level thresholding based approach. [11] proposed using a UNet combined with post processing methods to increase the segmentation accuracy. [12] explore different loss functions, and encoder pretraining.

Similarly, the field of bone supression has been explored for some time now. [13] attempt it using regression by k-nearest neighbor averaging, while [14] used multiple massive-trained artificial neural networks, a variant of fully connected networks. [15] suggests using Generative Adversarial Networks as a backbone for this task.

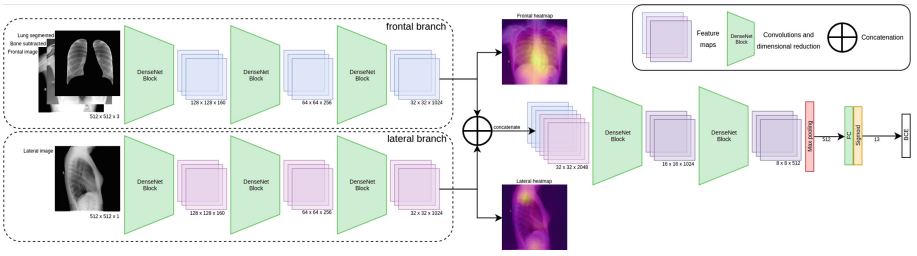


Fig. 2. Dual branch architecture

2 Our Contribution

In this work we create an end-to-end model for x-ray pathology classification based on the DenseNet-121 [16] architecture. We train it using custom augmentations on large public datasets, modify the network to support larger, 512×512 inputs and add bone subtracted and lung segmented versions of the inputs to increase accuracy. We propose a new dual branch architecture based on DenseNet-121 to increase accuracy in cases where lateral images are available.

The augmentation techniques we use to train all of our networks are based on [17]. We improve them by adding level adjustment, which is relevant for x-ray diagnostics and is used by radiologists in practice.

We add a stride 2 convolution and a stride 2 max pooling layer at the input of our DenseNet-121, to quickly reduce the dimensionality of our inputs from 512×512 to 128×128.

We trained 2 auxiliary models, a lung segmentation one (which increased our mean AUC by 0.004), and a bone subtraction one (which increases the average AUC by an additional 0.021). Coupled with our 2 branch architecture, this increases the mean AUC of our baseline by 0.038. Our current results are presented in Table 1.

3 Methods

3.1 Lung Segmentation

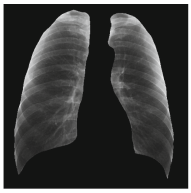
Lung segmentation is the task of extracting the lung area from an x-ray, removing areas that are considered irrelevant for diagnosing lung diseases. This can help radiologists better focus on the part of the image they are most interested in, as well as make it easier for them to take relevant measurements that they further use in the process of diagnostics.

Observing that lung segmentation would also reduce background noise for our pathology classification algorithm and help the model focus on more relevant areas, we set out to implement a model capable of segmentation (Fig. 1-c).

The model we train is built on a UNet [18] architecture. Instead of opting for the standard model described in the paper, we decide to use VGG16 [19] as the encoder and adapt the decoder to it, as [12] suggested. The main reason for this is that the ImageNet pretraining that can be applied to the VGG16 encoder improves the performance of the model. Unlike the author, we trained on both the Montgomery and the JSRT datasets, a total of 800 chest x-rays that were manually segmented by radiologists. We use data augmentation in the form of small rotations and translations. Our model achieves a Dice score of 0.96 on the validation dataset, consisting of 100 randomly sampled images.

One notable aspect of our training's results is that the model can achieve similar performance on datasets coming from different data distributions. Every x-ray generating machine generates radiographs slightly different, data coming from different machines varying in their distributions. Most of the CAD systems are not portable from one data distribution to another: the change makes them

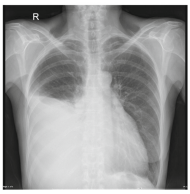
(a) Normal x-ray



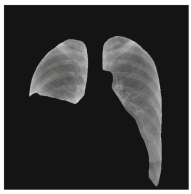
(b) Normal x-ray lung segmented

(c) Collapsed lung x-ray

(d) Collapsed lung x-ray lung segmented



(e) Partially collapsed lung x-ray



(f) Partially collapsed lung x-ray lung segmented

Fig. 3. Limitations of the current lung segmentation algorithm

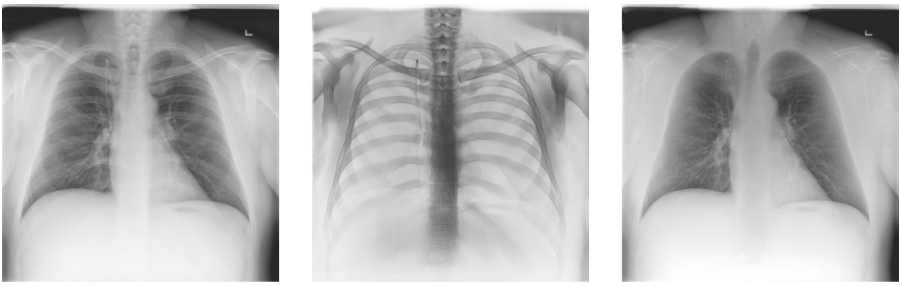
perform worse if the task they have to perform is non-trivial. The augmentation that we used, combined with the relative simplicity of the task and the use of 3 wildly different data distributions in the training dataset helped our model achieve a similar Dice score on other data.

A problem with this model, however, is that in case of lung obstruction, it will only partially segment the lungs. This can lead to downgraded performance down the line and is a good topic for further research. An example is shown in Fig. 3.

3.2 Bone Subtraction and Suppression

Bone subtracted x-rays are versions of the original images that strongly emphasize the skeletal tissue. They are useful in clinical settings as they help radiologists spot fractures and calcification signals, which can be hard to identify on the simple radiograph. They also present the advantage of being subtractive from the original radiograph, which leads to bone suppressed x-rays, in which the bone structure is not visible and the soft tissue can be seen better. Refer to Fig. 4.

The images are produced with the help of 'Dual Energy' enabled x-ray machines. Those expose the patient to two waves of radiation in an extremely fast sequence. The first wave has a higher dosage, while the second one is lower in intensity. This is the mechanism that allows the hardware to produce the bone subtracted image, and it is the most reliable way to obtain them (Fig. 1-b).



(a) Original x-ray (b) Bone Subtracted (c) Bone Supressed

Fig. 4. Bone subtraction and supression on an image from the ChestXRay-14 dataset

Our motivation for generating bone subtracted images with the help of artificial intelligence is that the method can yield an accuracy close to the one mentioned above while solving two major problems that standard hardware poses. The first one is that the second irradiation doesn't need to happen if using artificial intelligence, and the second is that 'Dual Energy' hardware is more expensive and not available as much as classic x-ray machines are.

To achieve the same effect as hardware techniques, we collect an anonymized dataset of 350 bone subtracted x-rays from the Oradea County Hospital. We use them to train a Conditional Deep Convolutional Generative Adversarial Network that can create a bone subtracted image. We start with the architecture of [15], and add [20]'s self attention, spectral normalization and update the discriminator and generator with the Two Times Update Rule (TTUR).

Doctors look after the contour of the bones when searching for fractures: thus, the resolution of the final image is extremely important. Our model benefits from adding a progressive multi-scaling strategy that helps it retain more of the exact structural information presented in the original radiograph.

After training our model and testing it on real cases, we conclude that the artificial intelligence, GAN technique can solve another problem that the hardware version encounters in practice: motion artifacts. The sequence of exposures of the 'Dual Energy' equipment is extremely fast, but the heart beating or slight movements of the patient can add artifacts to the image. While the full dataset that we have also contains artifact-broken images, we eliminate those and train on the 350 that contain little to no artifacts. The GAN model learns only from images without major artifacts and can generate artifact free bone subtracted images.

3.3 Pathology Classification

The most time consuming and important task of a radiologist with regards to x-ray analysis is diagnosing the study: identifying pathologies in the images at hand. Inspired by [3], we create an end-to-end classification model trained and validated on the official splits of the CheXPert and MIMIC-CXR datasets to help doctors with this difficult task.

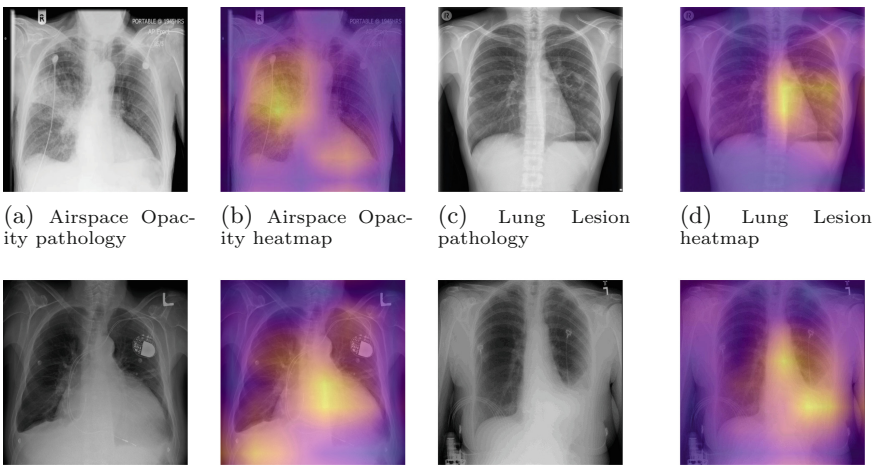
We test multiple architectures like ResNet, Inception, VGG, and DenseNet. We conclude that DenseNet works best for our task and decide to use a DenseNet-121 architecture with 512×512 RGB image inputs as a baseline. The choice for a larger input image size slows down the training of the network, but the decision was driven by multiple consulations with radiologists. They suggested that smaller images do not retain enough detail on incipient pathologies or on ones that are well hidden.

Debugging neural networks is a notoriously difficult task, and artificial intelligence algorithms applied to clinical tasks have to be consistently stable and well tested before used in real scenarios. To see that the algorithm correctly links diagnostics to their appearances in the radiography, we explain its classification results by using the CAM [21] visualization technique. Some results are shown in Fig. 5.

We conclude that using this visualization method has two other benefits besides helping with debugging. It strengthens the doctor's trust in the CAD system, as he can see that the diagnostics the software suggests have good reasoning behind them, and it also guides doctors, after they gain trust in the model, to faster navigate the important areas of a radiographs.

We evaluate the model using the AUROC score (see Table 1), as it is the most widely used for quantifying results in this task.

Our baseline starts with ImageNet dataset pretraining. Because x-rays are very different from the images in the Imagenet [22] dataset, we decide to abandon the ImageNet transfer learning used by [3]. Instead, we develop a binary



(a) Airspace Opacity pathology (b) Airspace Opacity heatmap (c) Lung Lesion pathology (d) Lung Lesion heatmap

(e) Cardiomegaly pathology (f) Cardiomegaly heatmap (g) Effusion pathology (h) Effusion heatmap

Fig. 5. Examples of pathologies and generated heatmaps

classification algorithm trained on the ChestX-ray14 dataset and use it to initialize the weights from our model. This greatly improves the speed at which the model reaches good performance and increases its AUROC score.

The binary classification model can tell normal radiographs apart from those that contain anomalies, without the ability to specify the exact anomaly that it encounters. While its main use was that of pretraining for our classification algorithm, we discuss possible use-cases with radiologists and reach the consensus that it is suitable for patient triage in highly dynamic and time-sensitive environments.

We train a binary classification algorithm and use it to initialize the weights of our model. The images resulting from our lung segmentation and GAN models provide valuable information to radiologists in real-world cases, so we decide to test if they help our model's performance as well.

To combine the images for the input, we switch to a single grayscale channel for the original x-ray image and add lung segmented and bone subtracted images (see Fig. 1) as inputs on separate channels.

Another tool used by radiologists in clinical scenarios is the lateral view radiographs. Some pathologies are not visible on frontal x-rays, while others can be hidden and can be seen more clearly with the help of a lateral investigation. With this in mind, we also add a second branch for a lateral chest x-ray, if one is available, believing that this will help the model associate anomalous areas in the frontal version with the ones present in the lateral version, for more accurate and comprehensive classification.

The lateral and frontal branches are constructed from 3 of the 5 Dense blocks normally used in DenseNet. Their outputs are firstly used to generate CAM visualization for both the images and are concatenated and fed into the last 2 Dense blocks, which provides enough depth for the network to learn correlations between the feature maps of the two branches.

Each of the images provides extra valuable information to the model, and adding them improves the model's performance. The architecture is described in Fig. 2.

4 Results

There are the results we obtained with different network and input configurations. There is a clear distinction in performance (as measured by the average AUC across all diseases) between them.

Table 1. AUC scores for various network architectures

Architecture	AUC score													
	AT	CA	CO	ED	PE	PN	PX	EC	LL	AO	PO	FR	SD	AVG
DN-121	0.827	0.813	0.827	0.892	0.922	0.716	0.812	0.795	0.675	0.752	0.822	0.494	0.849	0.784
DualBranch DN-121	0.843	0.831	0.837	0.910	0.934	0.733	0.817	0.814	0.683	0.763	0.828	0.500	0.866	0.797
+LungSegmentation	0.847	0.834	0.842	**0.913**	**0.938**	**0.735**	0.825	**0.817**	0.685	**0.773**	**0.833**	0.504	**0.871**	0.801
+BoneSubstracted	**0.851**	**0.841**	**0.845**	0.905	0.937	0.734	**0.827**	0.814	**0.693**	0.764	0.824	**0.784**	**0.871**	**0.822**

Classes: AT - Atelectasis; CA - Cardiomegaly; CO - Consolidation; ED - Edema; PE - Plerual Effusion; PN - Pneumonia; PX - Pneumothorax; EC - Enlarged Cardiomediastinum; LL - Lung Lesion; AO - Airspace Opacity; PO - Pleural Other; FR - Fracture; SD - Support Devices

DN-121 - baseline architecture, a DenseNet-121 with our initial downsampling layer, trained with augmented images

DualBranch DN-121 - our dual branch architecture, which takes 2 grayscale 1 channel inputs: frontal x-ray and lateral x-ray (if available)

+LungSegmentation - dual branch architecture, with an additional channel on the frontal input, which contains the lung segmented image

+BoneSubtracted - dual branch architecture + lung segmentation, with an additional channel on the frontal input, which contains bone subtracted images (see Fig. 2)

5 Discussion

Compared to other studies, by achieving a higher overall AUC we show that deep learning algorithms can benefit from learning with additional radiography information, which is natural because radiologists themselves find the auxiliary images useful.

Although we managed to increase our average AUC by combining different supplementary inputs, we were surprised to find that the AUC scores of some classes such as Edema and Pleural Effusion were decreased - this suggests that the dataset we are validating on has inconsistencies in how accurate the classes it contains are obtained.

This uncovers a major limitation of this study: the dataset used for training is labeled with NLP algorithms which provide an effective way of creating a large database of images and labels, but not an error-prone one. We believe a sharp rise in the algorithm performance could be provided by a radiologist hand-labeled dataset and we think that using such a dataset at least for validation is a necessity in future works.

6 Conclusion

In conclusion, we manage to improve the performance of our baseline model by increasing the size of the input, tweaking the DenseNet backbone, using weight initialization from a task performed on the same data distribution and by adding as inputs supplementary images, created by our 2 auxiliary models, that hold valuable information about the case. The lung segmentation model performs well on different data distributions than the ones it was trained on, while the GAN model that generates bone subtracted images from classic x-rays can solve a problem that consecrated hardware methods encounter: motion-artifacts.

In the future we plan to investigate what other auxiliary images can help the model learn even more detail about the x-ray it has to interpret, as well as to

improve the backbone architecture by adding attention modules and investigating multi-task learning approaches.

Lung Masks for Shenzhen Hospital Chest X-ray (Montgomery) Set:

- Dataset: https://www.kaggle.com/eduardomineo/u-net-lung-segmentation-montgomery-shenzhen/data
- Data Sources
  - National Library of Medicine, National Institutes of Health, Bethesda, MD, USA;
  - Computer Engineering Department, Faculty of Informatics and Computer Engineering, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine;
- Works using this dataset:
  - Jaeger S, Karargyris A, Candemir S, Folio L, Siegelman J, Callaghan F, Xue Z, Palaniappan K, Singh RK, Antani S, Thoma G, Wang YX, Lu PX, McDonald CJ. Automatic tuberculosis screening using chest radiographs. IEEE Trans Med Imaging 2014;33:233-45.
  - Candemir S, Jaeger S, Palaniappan K, Musco JP, Singh RK, Xue Z, Karargyris A, Antani S, Thoma G, McDonald CJ. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. IEEE Trans Med Imaging 2014;33:577-90.
  - Yu. Gordienko, Yu. Kochura, O. Alienin, O. Rokovyi, S. Stirenko, Peng Gang, Wei Zeng, Chest X-Ray Analysis of Tuberculosis by Deep Learning with Segmentation and Augmentation, arXiv preprint arXiv:1803.01199 (2018).

JSRT dataset: Japanese Society of Radiological Technology

- https://www.jsrt.or.jp/data/english/jsrt/.

References

1. PAHO WHO. World Radiography Day: Two-Thirds of the World's Population has no Access to Diagnostic Imaging. PAHO (2012)
2. Stec, N., Arje, D., Moody, A.R., Krupinski, E.A., Tyrrell, P.N.: A systematic review of fatigue in radiology: is it a problem? AJR Am. J. Roentgenol. **210**(4), 799–806 (2018)
3. Rajpurkar, P., et al.: CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. arXiv e-prints, art. arXiv:1711.05225, November 2017
4. Yao, L., Poblenz, E., Dagunts, D., Covington, B., Bernard, D., Lyman, K.: Learning to diagnose from scratch by exploiting dependencies among labels. arXiv e-prints, art. arXiv:1710.10501, October 2017
5. Kumar, P., Grewal, M., Srivastava, M.M.: Boosted Cascaded Convnets for Multilabel Classification of Thoracic Diseases in Chest Radiographs. arXiv e-prints, art. arXiv:1711.08760, November 2017
6. Guan, Q., Huang, Y., Zhong, Z., Zheng, Z., Zheng, L., Yang, Y.: Diagnose like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification. arXiv e-prints, art. arXiv:1801.09927, January 2018

7. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. arXiv e-prints, art. arXiv:1705.02315, May 2017
8. Irvin, J., et al.: CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. arXiv e-prints, art. arXiv:1901.07031, January 2019
9. Johnson, A.E.W., et al.: MIMIC-CXR: a large publicly available database of labeled chest radiographs. arXiv e-prints, art. arXiv:1901.07042, January 2019
10. Armato III, S.G., Giger, M.L., MacMahon, H.: Automated lung segmentation in digitized posteroanterior chest radiographs. Acad. Radiol. **5**(4), 245–255 (1998)
11. Rashid, R., Akram, M.U., Hassan, T.: Fully convolutional neural network for lungs segmentation from chest X-rays. In: Campilho, A., Karray, F., ter Haar Romeny, B. (eds.) ICIAR 2018. LNCS, vol. 10882, pp. 71–80. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93000-8_9. ISBN 978-3-319-93000-8
12. Frid-Adar, M., Ben-Cohen, A., Amer, R., Greenspan, H.: Improving the Segmentation of Anatomical Structures in Chest Radiographs using U-Net with an ImageNet Pre-trained Encoder. arXiv e-prints, art. arXiv:1810.02113, October 2018
13. Loog, M., van Ginneken, B., Schilham, A.M.R.: Filter learning: application to suppression of bony structures from chest radiographs. Med. Image Anal. **10**(6), 826–840 (2006). https://doi.org/10.1016/j.media.2006.06.002. http://www.sciencedirect.com/science/article/pii/S1361841506000454. ISSN 1361-8415
14. Chen, S., Suzuki, K.: Separation of bones from chest radiographs by means of anatomically specific multiple massive-training anns combined with total variation minimization smoothing. IEEE Trans. Med. Imaging **33**(2), 246–257 (2014). https://doi.org/10.1109/TMI.2013.2284016
15. Zhou, B., Lin, X., Eck, B., Hou, J., Wilson, D.: Generation of virtual dual energy images from standard single-shot radiographs using multi-scale and conditional adversarial network. In: Jawahar, C.V., Li, H., Mori, G., Schindler, K. (eds.) ACCV 2018. LNCS, vol. 11361, pp. 298–313. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20887-5_19
16. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely Connected Convolutional Networks. arXiv e-prints, art. arXiv:1608.06993, August 2016
17. He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., Li, M.: Bag of Tricks for Image Classification with Convolutional Neural Networks. arXiv e-prints, art. arXiv:1812.01187, December 2018
18. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv e-prints, art. arXiv:1505.04597, May 2015
19. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv e-prints, art. arXiv:1409.1556, September 2014
20. Zhang, H., Goodfellow, I., Metaxas, D., Odena, A.: Self-Attention Generative Adversarial Networks. arXiv e-prints, art. arXiv:1805.08318, May 2018
21. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning Deep Features for Discriminative Localization. arXiv e-prints, art. arXiv:1512.04150, December 2015
22. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: CVPR 2009 (2009)