American Journal of Neural Networks and Applications
2020; 6(2): 22-28
http://www.sciencepublishinggroup.com/j/ajnna
doi: 10.11648/j.ajnna.20200602.12
ISSN: 2469-7400 (Print); ISSN: 2469-7419 (Online)

Review Article

Overview of the Three-dimensional Convolutional Neural Networks Usage in Medical Computer-aided Diagnosis Systems

Bohdan Chapaliuk

Department of Mathematical Methods of Systems Analysis, Institute for Applied System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

Email address:
Bohdan.chapaliuk@gmail.com

To cite this article:
Bohdan Chapaliuk. Overview of the Three-Dimensional Convolutional Neural Networks Usage in Medical Computer-Aided Diagnosis Systems. *American Journal of Neural Networks and Applications.* Vol. 6, No. 2, 2020, pp. 22-28. doi: 10.11648/j.ajnna.20200602.12

Received: August 4, 2020; **Accepted**: August 17, 2020; **Published**: August 27, 2020

Abstract: Medical computer-aided diagnosis systems are essential applications that help doctors speed up, standardize, and improve disease prediction quality. Nevertheless, it is hard to implement a high-accuracy diagnosis system due to complex medical data structures that are hard to interpret even by an experienced radiologist, lack of the labeled data, and the high-resolution three-dimensional nature of the data. Meanwhile, modern deep learning methods achieved a significant breakthrough in various computer vision tasks. Thus, the same methods began to gain popularity in the community that works on the computer-aided systems implementation. Most modern diagnosis systems work with three-dimensional medical images that cannot be processed by traditional two-dimensional convolutional neural networks to get high enough prediction results. Hence, medical research introduced new methods that use three-dimensional neural networks to work with medical images. Even though these networks are usually an adapted version of state-of-the-art two-dimensional networks, they still have their specifics and modifications that help achieve human-level accuracy and should be considered separately. This article overviews the three-dimensional convolutional neural networks and how they are different from their two-dimensional versions. Moreover, the article examines the most influenced systems that achieve human-level accuracy in predicting the specific disease. The networks discussed in the perspective of two basic tasks: segmentation and classification. That is because the simple end-to-end classification neural networks usually do not work well on the available amount of data in the medical domain.

Keywords: Three-dimensional Convolutional Neural Networks, Medical Imaging, Deep Learning

1. Introduction

Lately, computer vision algorithms achieved a significant breakthrough through extensively applying deep learning approaches [1]. That led to the popularization and dissemination of deep learning methods in scientific and engineering communities. The basis of the breakthrough was using deep convolutional neural networks (CNN) that consist of many layers [2]. The wide use of CNN for optical information analysis attracted the attention of scientists who mainly work on creating medical applications that analyze medical images to predict patient diagnose in an automated way [3]. Applying deep learning methods to medical imaging allows achieving high-quality results, that in some tasks, rival the performance of an average human radiologist [4-7]. Analyzing medical images and data, however, remains quite a challenging task. There are several reasons for that: medical data contain quite complex structures that even humans find hard to interpret [8]; available datasets are generally small and have three-dimensional natures. In addition, medical data is hard to collect and annotate since it sometimes requires painful medical procedures and operates with sensitive private data that should not be publicly available or associated with any person. Thus, the building of more data-efficient and secure diagnosis systems is essential to maintain medical application prediction quality progress. Most of the medical data are three-dimensional images acquired based on computed tomography (CT) [9], magnetic resonance imaging

(MRI) [10], endoscopy video data [5], and so on. Hence, we believe that three-dimensional convolutional neural networks are among the deep learning models that can be used to improve current medical computer-aided diagnosis systems.

In the scope of this article, we will consider how three-dimensional convolutional neural networks can be utilized to analyze medical images and how they are already used to achieve human-level performance in disease prediction.

2. Convolutional Neural Networks

Convolutional neural networks (CNN) are usually defined as special kind neural networks that use mathematical convolution operation in at least one of the neural network layers. The first successful CNN was introduced in 1998 by Yann LeCun et al. [2] for the handwritten digits recognition task. Proposed architecture LeNet was able to work with the image represented as two-dimensional data and use two-dimensional spatial information to achieve state-of-the-art results for that time.

Convolutional neural networks were created with several architectural ideas in mind, such as local receptive fields that allows sparse interaction between layer activations; parameter sharing that allows reuse of neural network weights across different image locations; and spatial sub-sampling that allows networks to be invariant to small changes in the input data [2]. All those properties are incorporated into the structure of the neural network, which, in typical form, consists of several convolutional layers and a fully connected layer (multi-layer perceptron) at the end. The typical convolutional layer consists of three essential components: convolution operation, activation function, and a pooling layer, as shown in figure 1.

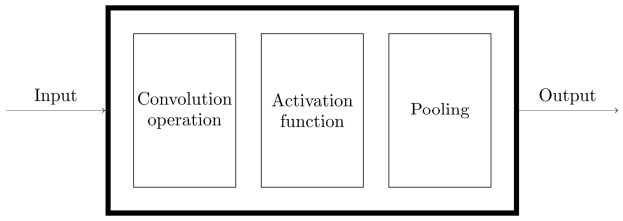

Figure 1. *Typical layer of the classic convolutional network. In modern network architectures, pooling stage might be omitted in some layers.*

For the two-dimensional data, such as image, convolution operation formula for image I and two-dimensional kernel K in a discrete form can be written as:

$$S_{i,j} = (I * K)_{i,j} = \sum_m \sum_n I_{m,n} K_{i-m,j-n}$$

Usually, modern convolutional neural networks are built with machine-learning frameworks such as TensorFlow or Pytorch. These frameworks implement convolution operation as a cross-correlation function that does almost the same mathematical operations, except kernel transposition at the end that flips the kernel [11, 12]. That is because kernel flipping is not usually necessary for the neural network, and without kernel flipping, convolution operation works faster [13]. Thus, for most modern convolutional neural network implementations convolution operation can be written as cross-correlation:

$$S_{i,j} = (I * K)_{i,j} = \sum_m \sum_n I_{i+m,j+n} K_{m,n}$$

There are several more differences between mathematical convolution operation and convolution operation used in the neural networks. First, neural network convolution operation supports "padding" and "stride" parameters that give a possibility to trick sliding windows step and the output size of the matrix. The "padding" parameter allows padding input image with zero values. The stride parameter specifies the window step size of convolution operation. In addition, convolution operation in neural networks can have multiple channels, which means several kernels can be applied to the same input in parallel. In case of input data contains more than one channel (RGB image or network intermediate layer output), convolution operation calculated as [13]:

$$S_{i,j,k} = \sum_{l,m,n} I_{l,j+m-1,k+n-1} K_{i,l,m,n}$$

where I is an input image that represents as a 3D array that consists of the $I_{i,j,k}$ elements that represent value in channel I, in the row j, and column k; K is a kernel that represented as a 4D array that consists of $K_{i,j,k,l}$ elements giving weights between a unit in channel I of the output and a unit in the channel j of the input, with an offset of k rows and l columns between the output and input units.

As we can see from the formula, if input data contains multiple channels, then a two-dimensional convolution will be the sum of all channel values multiplied by corresponding kernel values. Therefore, two-dimensional convolution operation cannot describe the spatial relationships in all three directions since it moves only in two directions by width and height. It causes the neural network to lose spatial information in the depth direction (see figure 2a). For the 2D color input images, it is not a problem since the neural network works with two-dimensional data. However, for the medical imaging problem, we often must work with three-dimensional data, and thus, they should be described and approximated in three directions. To make this possible, a three-dimensional convolution operation can be used.

Three-dimensional convolution operation has a filter depth dimension smaller than the input data depth, and thus, sliding window moves in all three directions as a cube (see figure 2b). Formally, three-dimensional convolution in a position x, y, z at i-th channel can be written as

$$S_{i,x,y,z} = \sum_l \sum_{p=0}^{P-1} \sum_{1=0}^{Q-1} \sum_{r=0}^{R-1} I_{l,x+p-1,y+q-1,z+r-1} K_{i,l,p,q,r}$$

where p, q, r are the coordinates of the three-dimensional kernel of convolution operation; K is the kernel; I is three-dimensional data on l channel.

The three-dimensional convolution behaves like the

two-dimensional convolution operation and applies similar architectural concepts to the input data, such as local receptive fields and shared weights. The only difference is the fact that three-dimensional convolution operation is applied to the cube and uses a smaller cube as a kernel; meanwhile, two-dimensional convolution operation just sums up all the values in the third dimension. The visual differences between them are shown in figure 2.

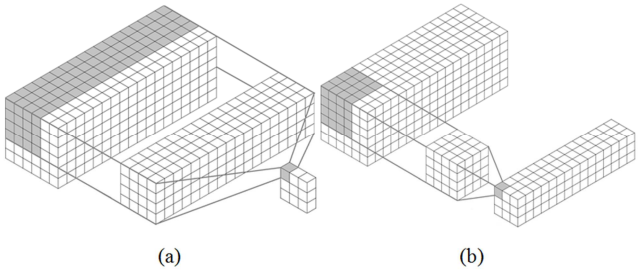

Figure 2. *These two images show the differences between 2D and 3D convolutions. Left side image shows how 2D convolution works with three-dimensional data. The right-side image shows that the 3D convolution applies cubic kernel to the three-dimensional data.*

For the first time, a three-dimensional convolution operation has been applied to human action recognition task in the video [14]. It was successful and outperformed other models that use recurrent neural networks to work with information from the temporal dimension in the video. The same approach can be applied to endoscopy video analysis.

The second component of the convolutional network layer is an activation function. This stage is used to add the nonlinearity to the layer to detect patterns on the input data. The most popular activation function in convolutional neural networks is the rectified linear unit [15].

The last stage of the convolutional neural network layer is a pooling function. In the context of convolutional neural networks, pooling function is the function that replaces output from the activation stage based on the summary statistic of the nearby values in the output. For example, there can be a max-pooling function that calculates maximum value in the input data within a rectangular neighborhood [16]. Also, there are plenty of other popular pooling functions that are used in convolutional neural networks such as the average pooling or L^2 norm of a rectangular neighborhood [17]. In all types of pooling, one of the core ideas of the pooling layer is to reduce the size of the data and improve feature representation invariance to the small variation in the input. However, the downside of this is that the pooling stage loses lots of useful information that might be useful to improve the final accuracy of the neural network. That is the reason why modern architectures do not use a pooling function on each convolutional layer.

In three-dimensional convolution neural networks, the subsampling stage (pooling) is applied in the same manner as in the two-dimensional convolutional neural networks, except that it works with three dimensions. It allows network to operate on cubes instead of squares.

3D convolution, activation functions and 3D pooling, as described above, allow us to build a three-dimensional convolutional neural network that can be applied to medical images. In the next section, we will discuss how such three-dimensional convolutional neural networks can be used to build disease detection systems.

3. Three-dimensional Convolutional Neural Network Usage

In this section, we consider how three-dimensional neural networks can help build medical computer-aided diagnosis systems. First, we will describe the typical medical data that can be analyzed by three-dimensional neural networks. Then, describe existing systems and how they use 3D CNN to improve the results of different tasks.

3.1. Typical Medical Images and Computer-aided Diagnosis Systems

Computer-aided diagnosis systems are the systems that assist doctors with medical image interpretation to speed up and improve the quality of patient disease diagnosis. Usually, x-ray images, computed tomography (CT) [9], magnetic resonance imaging (MRI) [10], or endoscopy video data [18] can be used. All these images aim to visualize human internal organ structures, so that the doctors can diagnose patient disease and prescribe related treatments.

There are several types of computer-aided diagnosis systems that can be used in hospitals [19]:

1) Systems that detect and label potentially suspicious areas on a medical image. These areas should show the anomaly that causes a patient's disease. In this case, the main task is to reduce the load on the radiologist via automated detection and description of related areas.
2) Computer-aided systems that diagnose patient disease using available medical images. In the ideal case, the system should return a correct diagnosis without involving a radiologist. In practice, diagnosis systems still require validation from the experts, so it should also include information about the location of abnormal areas.

The actual implementation of an automated diagnostic system heavily depends on the input data that is used for making diagnoses because it might require different preprocessing steps. 3D convolutional neural networks are usually used for CT scans, MRI, and endoscopy video data.

For example, modern medicine practice recommends using a low-dose CT scan of the human chest to investigate lung cancer presence [20]. CT scan represents a three-dimensional image of the patient's lungs obtained by the X-ray that gradually passes through the human body's tissue, layer by layer, in different directions, angles, and positions. Combined layers of such images form a 3D image of the patient's chest that can be used as an input to the 3D convolutional neural network. The same can be done for other body parts. For each distinct disease, a separate dataset should be created, and specialized network trained.

Magnetic resonance imaging is another type of the layer-by-layer sequence scan that uses strong magnetic fields

and radio waves to generate image of the organ's internal representation in the human body. It enables getting the same slice-by-slice image of the required human part and combining all of them to get a 3D image that can be analyzed.

3.2. Building Medical Applications that Analyze Medical Images

In the simple case, to build the computer-aided diagnosis system that works with a CT scan or MRI, we can reuse a simple 3D convolutional neural network that gets 3D images as input and makes a prediction as output. For example, it is possible to reuse convolutional neural networks like C3D [21], 3D ResNet [22], or 3D DenseNet [23]. 3D ResNet and 3D DenseNet neural networks are an adapted version of the ResNet and DenseNet that are used for two-dimensional image classification and were initially used on the ImageNet data. An example of the 3D DenseNet architecture that can be used for patient disease prediction is shown in figure 3.

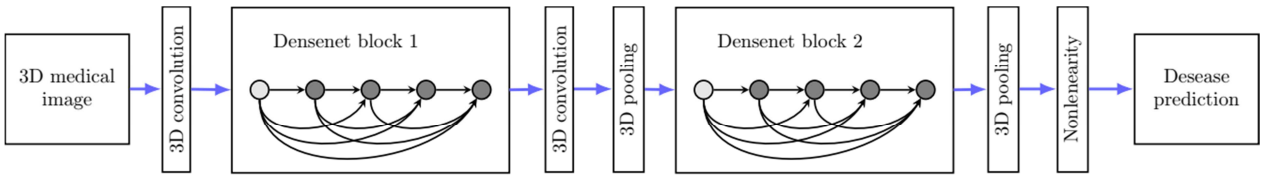

Figure 3. *An example of the 3D DenseNet for the patient disease prediction based on the three-dimensional medical data like CT scans or MRIs.*

In addition, any modern convolutional neural network architecture, that shows state-of-the-art results on the ImageNet dataset, can be adapted to the 3D data by replacing two-dimensional convolution and pooling operations with its three-dimensional version. However, the downside of the approach is lack of the ability to use transfer learning to speed up the neural network training.

In practice, using a 3D convolutional neural network for classification does not work well since network training requires a vast amount of the data that is a rare case in medical domain. For example, such neural network training results show accuracy only of 70%-72% for the lung cancer detection problem [24]. For the modern computer-aided diagnosis systems, this is not enough.

To improve the diagnosis system's prediction results, the diagnosis process can be divided into two smaller tasks: segmentation and classification. In the segmentation task, the neural network should find any possible abnormal lesion region in the medical image. Then, the classification step model should consume the abnormal region at the segmentation stage and predict the presence of the disease in the patient [4, 6]. Let us consider each task in more detail.

3.2.1. Segmentation Task

The main goal of the segmentation task is to assign labels to each image pixel so that each region and object on the image becomes associated with a corresponding label. The typical convolutional neural network architecture that is used for analyzing 2D medical imaging is U-Net [25]. U-Net is a fully convolutional neural network that consists of two parts: a contracting path that follows typical convolutional network architecture described in the previous section and an expansive path that does upsampling of the feature maps from the contractive path. The peculiarity of this architecture is the use of feature maps from corresponding contractive path intermediate layer in the corresponding intermediate layer of the expansive path. This technique helps solve the problem of losing information when applying convolution operation due to the decreased size of the input data. U-Net showed state-of-the-art results on the ISBI cell tracking challenge and high accuracy, even on the small datasets. Soon, this architecture was used and adapted to the different tasks in the medical domain [26, 27].

The U-Net architecture was adapted to the three-dimensional medical images as a 3D U-Net [28-30] or V-Net [31] network. To adapt U-Net to the three-dimensional data, all convolutional operations were replaced by three-dimensional convolutions. In the expansive path, all upsampling operations were replaced by three-dimensional upsampling operations. That solution works quite well and allows for high results on different tasks. For example, work [31] reports an average dice coefficient of 0.869 on prostate MRI segmentation task. Likewise, applying three-dimensional U-Net architecture to the brain tumor segmentation problem [32] allows achieving a dice coefficient of 0.858. In the last years, this fully convolutional network architecture is still popular; however, it is applied with modifications that allow building more deep architecture [26].

There are several other convolutional network architectures that can help with region detection in the medical images; however, they were initially used for object detection outside the medical domain. First, it is the Faster R-CNN [33] network that was a winner of COCO 2015 and ILSVRC 2015 object detection competitions. This detector basically consists of two stages. In the first stage, the Region Proposal Network (RPN) generates a possible object bounding box. Then, in the second stage, another neural network extracts feature maps from each generated possible object bounding box and performs regression and classification. As a result, Faster R-CNN returns the objects bounding box and labels them as associated with corresponding supported objects.

Faster RCNN architecture can be adapted to the three-dimensional data by replacing two-dimensional convolution operations with three-dimensional versions. For example, 3D Faster RCNN detectors were used in the work [34] for multi-organ segmentation in head and neck on MRI images, and were able to achieve a dice coefficient of 0.89 in the best case.

Finally, for the instance segmentation of the regions in the images can be used the Mask RCNN [35]. It is a two-stage convolutional neural network that modifies Faster RCNN architecture by adding binary mask output for each generated possible object bounding boxes in the second stage. This approach can be adapted to the three-dimensional medical data in the same way as Faster RCNN. For example, 3D Mask RCNN was used for the kidney and tumor segmentation tasks [36], and pulmonary nodule detection problems [37]. There were dice similarity coefficients of 0.96 and 0.80 reported for each task, respectively. In addition, 3D Mask RCNN was an important part of the four-stage neural network for the lung cancer screening system [4] that achieved the accuracy level of the average radiologist.

3.2.2. Classification Task

Classification of the diseases based on the medical images using the deep learning approach gained popularity lately. Usually, classification tasks are used in combination with a segmentation task, where the abnormal regions for classification are obtained by the segmentation task [4, 6]. This approach gives the radiologist an advantage of more transparent work of computer-aided diagnosis system because it gives information about why the system predicts disease. In addition, doing classification based on the segmented regions gives a possibility to reduce memory consumption and usage of computational resources on the classification stage.

The classification works with three-dimensional regions (cubes) that were segmented on the segmentation stage. Indeed, the classification model should reason about disease presence using some portions of the segmented regions. It can be implemented by classifying all generated regions with the classification model and then combining results with the multi-instance learning assumption. This assumption says that if there exists an instance that is positive, the whole bag is positive. The whole bag is negative if all instances are negative [38]. Thus, the classification network should be run on all segmented regions, and if some of them give a high probability of the disease, the system should conclude positively about disease presence.

The three-dimensional segmented regions usually processed by the three-dimensional classification neural networks. These networks typically represent two-dimensional neural networks, initially created for ImageNet, with adaptation to the three-dimensional data. For example, it might be adapted architectures like ResNet [39] or DenseNet [40]. As with the segmentation task, adaptation to three-dimensional data usually can be made by replacing existing convolution and pooling operations with three-dimensional counterparts. Such networks help achieve high results in different domains [41, 42]. For example, in the lung cancer screening application, combining segmentation and classification task gives an accuracy of 95% [4].

There are research papers that show results using the 3D CNN for classification only. However, as discussed earlier, that method does not show good results and requires a vast amount of data to train the network [24, 43].

4. Conclusion

Most modern medical computer-aided diagnosis systems that work with three-dimensional medical images or video data use 3D convolutional neural networks under the hood. Usually, 3D CNN architecture mimics the known two-dimensional CNNs; however, they replace all the convolution and pooling operations with their three-dimensional counterparts. Moreover, to achieve human-level accuracy in disease prediction, the system splits prediction into the smaller tasks of segmentation and classification. In the segmentation stage, the network should find three-dimensional regions that contain probable abnormal regions; meanwhile, classification three-dimensional neural networks use segmentation outputs and multi-instance learning assumption to evaluate the probability of disease in the image. As shown in the paper, each step allows the building of diagnosis systems with a human-level accuracy in many medical domains.

Nevertheless, pipelines with three-dimensional (3D) CNNs have their drawbacks. First, 3D convolution operation requires much more computation and memory resources in comparison to the two-dimensional counterpart. In fact, 3D convolution operation resource requirements grow cubically with increasing input size limiting the input medical image size due to GPU memory constraints.

Second, there are no pre-trained 3D convolutional neural networks that can be used for transfer learning. Therefore, the networks should be trained from scratch. However, this issue can be easily solved as far as research goes, in case the solutions will be publicly available.

Third, existing state-of-the-art pipelines still require a vast amount of data to achieve high disease prediction results and require much data-labeling work. This is a problem for the medical domain since, usually, not much data are available, and radiologists' work is expensive. Therefore, modern systems should still consider improving network-training efficiency.

Finally, modern systems lack the interpretability that might cause diffidence on the system by radiologists. It is a common problem for all the neural networks since they come as a black-box model that improves its work via consumed data on the training stage. Therefore, tools and methods that help with neural network system interpretability should still be created. Indeed, it will help medical applications and make it possible to build trust between humans and automated systems.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017, doi: 10.1145/3065386.

[2] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, 1998, doi: 10.1109/5.726791.

[3] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift fur Medizinische Physik*. 2019, doi: 10.1016/j.zemedi.2018.11.002.

[4] D. Ardila *et al.*, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," *Nat. Med.*, 2019, doi: 10.1038/s41591-019-0447-x.

[5] T. Ross *et al.*, "Exploiting the potential of unlabeled endoscopic video data with self-supervised learning," *Int. J. Comput. Assist. Radiol. Surg.*, 2018, doi: 10.1007/s11548-018-1772-0.

[6] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, "Evaluate the Malignancy of Pulmonary Nodules Using the 3D Deep Leaky Noisy-or Network," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3484-3495, Nov. 2019, doi: 10.1109/TNNLS.2019.2892409.

[7] W. Zhu, C. Liu, W. Fan, and X. Xie, "DeepLung: Deep 3D dual path nets for automated pulmonary nodule detection and classification," *Proc. - 2018 IEEE Winter Conf. Appl. Comput. Vision, WACV 2018*, vol. 2018-Janua, pp. 673–681, 2018, doi: 10.1109/WACV.2018.00079.

[8] S. G. Armato *et al.*, "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans," *Med. Phys.*, 2011, doi: 10.1118/1.3528204.

[9] J. A. Neutze, "Computed tomography," in *Radiology Fundamentals: Introduction to Imaging & Technology*, 2020.

[10] R. W. Chan, J. Y. C. Lau, W. W. Lam, and A. Z. Lau, "Magnetic resonance imaging," in *Encyclopedia of Biomedical Engineering*, 2018.

[11] Tensorflow, "tf.nn.convolution." https://www.tensorflow.org/api_docs/python/tf/nn/convolution.

[12] Pytorch, "Convolution layers." https://pytorch.org/docs/stable/nn.html#convolution-layers.

[13] A. C. Ian Goodfellow, Yoshua Bengio, "Deep Learning Book," *Deep Learn.*, 2015, doi: 10.1016/B978-0-12-391420-0.09987-X.

[14] S. Ji, W. Xu, M. Yang, and K. Yu, "3D Convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, 2013, doi: 10.1109/TPAMI.2012.59.

[15] G. E. Hinton and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair," Accessed: Mar. 13, 2020. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.165.6419.

[16] Y. T. Zhou and R. Chellappa, "Computation of optical flow using a neural network," 1988, doi: 10.1109/icnn.1988.23914.

[17] Y. L. Boureau, J. Ponce, and Y. Lecun, "A theoretical analysis of feature pooling in visual recognition," 2010.

[18] N. Lagattolla, "Endoscopy," in *Key Topics in General Surgery*, 2002.

[19] R. Takahashi and Y. Kajikawa, "Computer-aided diagnosis: A survey with bibliometric analysis," *International Journal of Medical Informatics*. 2017, doi: 10.1016/j.ijmedinf.2017.02.004.

[20] A. Neroladaki, D. Botsikas, S. Boudabbous, C. D. Becker, and X. Montet, "Computed tomography of the chest with model-based iterative reconstruction using a radiation exposure similar to chest X-ray examination: Preliminary observations," *Eur. Radiol.*, vol. 23, no. 2, pp. 360–366, 2013, doi: 10.1007/s00330-012-2627-7.

[21] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 4489–4497, 2015, doi: 10.1109/ICCV.2015.510.

[22] A. Q. and C. K. A. and S. S. and M. K. A. A. K. and L. W. H. and M. M. and T. D. Chung, "Hybrid 3D-ResNet Deep Learning Model for Automatic Segmentation of Thoracic Organs at Risk in CT Images," *2020 Int. Conf. Ind. Eng. Appl. Manuf.*, pp. 1–5, 2020.

[23] T. D. Bui, J. Shin, and T. Moon, "3D Densely Convolutional Networks for Volumetric Segmentation," 2017, [Online]. Available: http://arxiv.org/abs/1709.03199.

[24] B. Chapaliuk and Y. Zaychenko, "End-to-End Deep Learning Strategies for Computer-Aided Lung Cancer Detection Systems," vol. 4, no. 5, pp. 140–155, 2019.

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, May 2015, vol. 9351, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[26] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," 2018, doi: arXiv: 1807.10165v1.

[27] P. F. Jaeger *et al.*, "Retina U-Net: Embarrassingly Simple Exploitation of Segmentation Supervision for Medical Object Detection," Nov. 2018, Accessed: Apr. 17, 2020. [Online]. Available: http://arxiv.org/abs/1811.08661.

[28] O. Cicek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," *Med. Image Comput. Comput. Interv. - MICCAI 2016*, 2016, doi: 10.1007/978-3-319-46723-8.

[29] H. Hwang, H. Z. Ur Rehman, and S. Lee, "3D U-Net for skull stripping in brain MRI," *Appl. Sci.*, 2019, doi: 10.3390/app9030569.

[30] S. Peng, W. Chen, J. Sun, and B. Liu, "Multi-Scale 3D U-Nets: An approach to automatic segmentation of brain tumor," *Int. J. Imaging Syst. Technol.*, vol. 30, no. 1, pp. 5–17, 2020, doi: 10.1002/ima.22368.

[31] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," *Proc. - 2016 4th Int. Conf. 3D Vision, 3DV 2016*, pp. 565–571, 2016, doi: 10.1109/3DV.2016.79.

[32] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10670 LNCS, pp. 287–297, Feb. 2018, Accessed: Aug. 01, 2020. [Online]. Available: http://arxiv.org/abs/1802.10508.

[33] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[34] Y. L. and J. Z. and X. D. and T. W. and H. M. and M. W. M. and W. J. C. and T. M. L. and X. Yang, "Multi-organ segmentation in head and neck MRI using U-Faster-RCNN," in *Medical Imaging: Image Processing*, 2020.

[35] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-Octob, pp. 2980–2988, 2017, doi: 10.1109/ICCV.2017.322.

[36] C.-Y. C. and L. M. and Y. J. and P. Zuo, "Kidney and Tumor Segmentation Using Modified 3D Mask RCNN," 2019.

[37] E. K. and G. Engelhard, "Lung Nodules Detection and Segmentation Using 3D Mask-RCNN," *ArXiv*, vol. abs/1907.0, 2019.

[38] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," *Artif. Intell.*, vol. 89, no. 1–2, pp. 31–71, 2002, doi: 10.1016/s0004-3702(96) 00034-3.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Dec. 2016, vol. 2016-Decem, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[40] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 2261–2269, 2017, doi: 10.1109/CVPR.2017.243.

[41] J. Zhou *et al.*, "Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in MR images," *J. Magn. Reson. Imaging*, 2019, doi: 10.1002/jmri.26721.

[42] K. O. and Y. C. and K. W. K. and W. K. and I. Oh, "Classification and Visualization of Alzheimer's Disease using Volumetric Convolutional Neural Network and Transfer Learning," *Sci. Rep.*, vol. 9, 2019.

[43] V. W. and S. A. and J. M. Buhmann, "Classification of brain MRI with big data and deep 3D convolutional neural networks," in *Medical Imaging*, 2018.