UNIVERSITY

FACULTY

STUDY PROGRAMME:

MASTER THESIS

Catchy Title

Submitted by:
John DOE
Matriculation Number: 123456
Street 123
789654 City

Initial Examiner:
Dr. Supervisor
Secondary Examiner:
Prof. Dr. Examiner

October 3, 2023

"Programming today is a race between software engineers striving to build bigger and better idiot-proof programs, and the universe trying to build bigger and better idiots. So far, the universe is winning."

Rick Cook

UNIVERSITY

Abstract

Faculty

Master Thesis

Catchy Title

by John DOE

Write your abstract here.

Keywords: LaTeX, Master Thesis, Programming

Acknowledgements

Ich bedanke mich bei Dr. Supervisor für die Betreuung dieser Arbeit.

Contents

List of Figures

List of Tables

Listings

List of Abbreviations

1 Introduction

Medical image segmentation, the process of delineating one region of the image such as cancerous tissues from the rest of the image, is a crucial step in computer-assisted medical image analysis. Whether the ultimate goal is surgical planning [1], diagnosis [2], performing measurements [3], or doing population-level research [4], segmentation is often the first step in understanding 2D and 3D medical images.

Neural networks have become the standard tool to achieve biomedical image segmentation in almost all problem areas including, among others, segmenting organs or specific tissues from CT, MRI, or X-ray images [5]; cells from microscopic images [6]; and skin lesions from dermatoscopic images [7].

While achieving impressive results, these results are highly dependent on the quantity and quality of the training data [8]. However, obtaining medical imaging data is challenging due to several reasons. Firstly, capturing medical images is costly both in terms of time and finances. For instance, MRI and CT scanning can take 30 minutes or more and require equipment that is inaccessible to large parts of the world. Secondly, such data is often large in terms of file size and stored in complex systems, increasing the friction of sharing and using the data. Finally, some jurisdictions define medical images as personally identifiable data [9] and thus require explicit consent for their secondary use for the purpose of training neural networks. While valid and understandable, these patient privacy concerns can limit the use of already existing large databases in medical institutions.

After obtaining a medical image, these images need to be labeled with high-quality delineations of the target region. Such labeling is often done through a tedious and time-consuming process where multiple experts would manually draw curves or polygons on the image. While some labeling methods make this process easier [10], each image still needs to be checked by an expert in the field. For challenging problems, this often requires highly trained and experienced specialists.

These challenges make collecting large medical image segmentation datasets infeasible. Therefore, to improve performance and robustness, there is a need to develop data-efficient segmentation methods. Data efficiency, sometimes referred to as sample efficiency, measures how well a model performs with respect to its sample size. Data-efficient models achieve good results given a small amount of data.

This thesis provides an overview of medical image segmentation and efforts to increase its data efficiency. It also presents several novel methods for achieving data efficiency in various contexts. The unifying central principle of these methods is the notion that necessary network capacity (number of parameters) grows with problem complexity. However, higher-capacity networks require a larger amount of data to be trained. The thesis aims to answer the question of how can we transform the data to make the segmentation problem simpler. This would allow us to train networks of lower capacity, and thus ones that are more data efficient. This is achieved using traditional image processing techniques informed by features detected by a neural network, thus combining the two worlds of traditional and neural network-based medical image segmentation.

1.1 Contributions of this thesis

This thesis aims to develop new methods of achieving data efficiency in medical image segmentation. In particular, we propose the following original contributions to the scientific literature:

1. A new biomedical image segmentation method based on polar transform preprocessing with a learned center point.
2. An improved method of reducing input image size in neural networks using salient image crops.
3. A new neural network architecture for high-resolution image segmentation that combines object detection in low-resolution images and segmentation in high-resolution images.
4. A new method of embedding depth information in two-dimensional convolutional neural network input data.

1.2 Organization of the thesis

todo

2 Neural Network-Based Segmentation of Biomedical Images

This chapter introduces biomedical images as well as the process of segmenting them. We will start with a brief overview of the most common types of biomedical images.

Biomedical images are a broad and diverse category of images referring to imaging use for the purposes of biology or medicine. These can range from complex modalities such as CT scans all the way to photographs (such as clinical images or photographs of plants). With such a diverse set of modalities and imaging techniques, biomedical images have a low level of cross-domain consistency. Methods developed for one modality often cannot be used in a different modality without modification. This is especially true for learned models such as neural networks. Biomedical image segmentation networks are notoriously bad at out-of-sample performance, i.e. when evaluated on a different modality than they were trained on. This chapter presents a brief overview of various biomedical imaging modalities and techniques, with a focus on those relevant to the methods presented in this thesis.

2.1 Common Types of Biomedical Images

Broadly, we may classify biomedical images into 3D and 2D images. 3D images include modalities such as CT and MRI, but also some types of microscopy as well as other techniques. 2D images, among others, include X-rays, most microscopic images, and dermatoscopic images. The images could also be classified by color space. Techniques that visualize the insides of objects such as CT, MRI or X-ray are usually grayscale. On the other hand, camera-based imaging techniques such as dermoscopy or photographs use three or more color channels. As is apparent, biomedical images do not yield themselves to clean classification, so the classification here is only a general descriptive categorization. There are novel techniques that combine different modalities that escape this classification.

2.1.1 3D Modalities

In medicine, 3D imaging of the patient allows experts to look below the patient's skin. It is hard to overstate the importance of modalities such as CT and MRI on modern medical developments and patient outcomes. These modalities are often captured and stored as voxel-based 3D files, where a voxel is the smallest 3D unit equivalent to a pixel in 3D. They are usually stored alongside patient data such as age, sex, imaging parameters and other relevant information.

Computed Tomography (CT)

Computed tomography is an x-ray-based imaging technique where different planes of the subject are captured and then reconstructed into a 3D image using a process called tomography. CT, as opposed to MRI, uses ionizing radiation that can be harmful in large

doses. This makes it important to evaluate the cost and benefit of each scan since each scan presents a risk of harming the subject. In addition, image quality is correlated with radiation dose, and the dose is carefully selected based on the required level of image quality so as to not unnecessarily radiate the subject.

Low-dose and high-dose CT images differ enough to cause issues in the segmentation model's performance across these two domains . High-quality images result in cite better neural network models, but their availability is much more limited than low-dose scans.

CT can be enhanced using a contrast agent. The contrast agent is usually injected intravenously and is chosen to appear with a high intensity on the resulting image. This makes it easier to delineate blood vessels from surrounding tissue, as is the case e.g. in angiography (CTA).

In the image itself, the values of a CT scan are usually stored as $W \times H \times D$ values, where W, H, and D are the x, y, and z dimensions, respectively. Each value represents a voxel, *volumetric pixel*, in a 3D volume. The voxels are usually not cubes and can have different lengths along each dimension. In the z-axis, the voxel length is known as slice thickness, and it is chosen based on the task at hand. Thinner slices increase the spatial detail in the image, leading to more details in small tissues. However, thinner slices generally also increase the level of noise in the image, so slice thickness is selected to maintain a tradeoff between spatial resolution and level of noise. The resolution along the x- and y-axes is governed by the field of view and the scanner itself.

The values of the voxels are usually stored as 12 bit signed integer values. The value corresponds to the attenuation of the X-ray as it passes through the tissue. Denser structures such as bones attenuate the radiation more strongly than adipose tissue, and thus result in higher values. This results in higher-intensity voxels in the resulting CT image. To maintain consistency across scans and machines, the attenuation values are linearly transformed such that air has a value of -1000 while water has a value of 0 at standard pressure temperature, as:

$$\mathrm{HU}(\mu) = 1000 \cdot \frac{\mu - \mu_{\text{water}}}{\mu_{\text{water}} - \mu_{\text{air}}} \tag{2.1}$$

where μ is the attenuation value of a voxel, and μ_{water} and μ_{air} are attenuation values of water and air, respectively. This transformation is called the Hounsfield unit (HU) scale. When calibrated in this way, the attenuation of each tissue is represented as relative to the attenuation of water, and thus values are standardized across images, CT machines, imaging parameters, and centers. Various tissues broadly fall into Hounsfield unit values as is shown in Table **??**.

Tissue	Hounsfield value
Fat	-30 to -70 HU
Muscle	20 to 40 HU
Bone	1000 HU
Blood	13 to 50 HU

Table 2.1: Approximate Hounsfield values of various tissues [11], [12].

Hounsfield values are a crucial tool in analyzing medical images. Aside from getting a better understanding of what tissues or objects are present, measured Hounsfield values can have diagnostic significance. For instance, clotted blood has a higher HU value than unclotted blood, and so the Hounsfield value is used as an indicator of intracranial

hemorrhage [12]. Another example is using the average HU value of epicardial fat as a sign of myocardial infarction [13].

Since the human eye recognizes much fewer gray levels than are available on a CT scan, the scan is typically windowed when viewed by an expert. Windowing refers to the process of shrinking the value range using two thresholds (t_1, t_2) as:

$$y(x) = \begin{cases} t_1, & \text{if } x \leq t_1 \\ x, & \text{if } t_1 < x < t_2 \\ t_2, & \text{if } x \geq t_2 \end{cases} \tag{2.2}$$

where x is a given HU value. This allows the software to visually stretch the remaining range of gray values to more easily see smaller differences in attenuation, as seen in figure **??**. This technique is not limited to human experts. It is common to window CT image inputs into segmentation neural networks. For instance, if segmenting fat, it is often beneficial to discard all voxels outside of the fat tissue range [14].

windowing figure

Magnetic Resonance Imaging (MRI)

Much like CT, MRI visualizes the insides of an object using a voxel-based image. MRI works by first applying a strong magnetic field such that protons align parallel to the z-axis. The protons are then excited using a radio frequency pulse which causes them to become misaligned. After the pulse, the protons gradually return back to alignment and induce an electric current. Unlike CT which measures the attention of X-rays, MRI measures this induced current while protons are returning to equilibrium alignment. The time to reach the equilibrium state depends on the specific tissue type. The use of magnetic fields instead of ionizing radiation means that MRI does not cause harm to the patient due to radiation. Unlike CT, MRI values are not standardized across scans and MRI machines, and the specific values can not be used diagnostically across scans, only in relation to surrounding tissues.

MRI generally provides a better contrast than CT, especially in soft tissues. This makes MRI the gold standard of imaging for a large number of tissues and diagnostic procedures. Much like CT, this contrast can be further enhanced using contrast agents. However, MRI imaging is slower when compared to CT which causes patient discomfort, especially for claustrophobic and non-neurotypical patients. MRIs are also not possible in cases where the subject has non-removable magnetic objects such as coronary pacemakers and other implants.

The voxels in an MRI, much like a CT, are generally rectangular solids and can different in the x-, y- and z-axis lengths. The resolution is government by the field of view, the MRI scanner as well as imaging parameters. Generally, increasing resolution leads to higher levels of noise and a longer acquisition time. Thus, a compromise needs to be determined for each imaged tissue and diagnostic task.

The voxel values are weighted based on the time to reach the equilibrium state which is achieved through two independent processes called T1 and T2. T1 measures the time it takes the protons to reach equilibrium longitudinal magnetization, while T2 measures the time to regain its equilibrium transverse magnetization. Different tissues regain equilibrium states in T2 and T2 at different rates, so MRI images can be weighted according to T1 or T2 time. Water has a long T1 time while fat has a short T1 time, so in T1-weighted images fat appears with a higher intensity than water. On the other hand, water has a high T2 and appears as high-intensity on T2-weighted images. This can be seen in figure **??**.

figure

The relative differences in T1 and T2 images make the two weightings more or less beneficial for imaging certain tissues. T1 weighting is useful, among others, for identifying fatty tissue or detecting liver lesions. T2 weighting is useful, among others, for identifying white matter lesions or edemas.

2.1.2 2D Modalities

While 3D modalities offer a detailed view of a subject, they are often time-consuming and invasive in the case of CT. 2D modalities such as X-ray and diagnostic ultrasound are often quicker and more available, making them ideal for screening and simpler diagnostics. The same reason also results in a much larger amount of publicly available datasets in 2D modalities compared to 3D ones. Segmentation, object detection and classification in X-ray images, for instance, is one of the most active fields in computerized medical image analysis research [15], [16].

X-ray imaging

Similarly to CT, X-ray imaging or radiography uses ionizing radiation emitted on one side of the object and detected on the other. The intensity of the image corresponds to the attenuation of the emitted radiation. Denser materials appear with a higher intensity on the resulting image due to their high attenuation.

When capturing an X-ray image, an expert manually positions the generator. The position of the generator and the object determine the magnification and field of view in the image. If the distance between the detector and the object is larger than the distance between the generator and the object, the object will appear larger on the resulting image. This magnification means that the scale on an X-ray image is relative to the image itself and so objective measurements cannot be made on an X-ray.

The expert also determines various parameters that ultimately change the quality and quantity of the X-ray beams. The quality measures the ability of an X-ray to penetrate tissue and is proportional to the X-ray energy level. Quantity, on the other hand, measures the number of photons that constitute the beam. Increasing the beam quality allows for imaging denser tissues such as bones or through large bodies but results in lower contrast in soft tissues. This means that intensity in X-rays is not standardized across images like it is in CT, and so there is no objective unit of measuring X-ray intensity.

Dermatological images (clinical and dermatoscopic)

An increasingly active application of deep learning-based models is in dermatological applications, namely skin lesion analysis. This is driven in part by organisations such as the International Skin Imaging Collaboration (ISIC) that curates large dermatological datasets [17]. These datasets consists of clinical and dermatoscopic images. Clinical images are regular photographs of skin lesions, while dermatoscopic images are captured with a digital epiluminescence dermatoscope. This dermatoscope consists of a camera attached to a magnifying lens with a built-in light source. It allows capturing detailed and magnified images of a skin lesion while filtering out skin reflections.

The primary application of deep learning in dermatological images is for classification — predicting whether a lesion is benign or not or detecting the type of skin disease. However, skin lesion segmentation plays a role in the detection as delineating the skin lesion border can be used to provide more objective attributes of the skin lesion [17].

Microscopy

In biomedicine, one of the most used applications of computer vision is in segmenting, analyzing, and quantifying microscopic images. This encompasses various tasks in digital pathology such as detecting cancerous cells, segmenting nuclei, or counting white blood cells.

Publicly available microscopic images of cells are abundant due to how frequently they are captured and that they don't contain any personally identifiable information. However, these images sometimes present a challenge due to their size. Microscopic images can have an area of multiple megapixels, far too large for current deep-learning models. Therefore, the images are often split into patches, downscaled or analyzed in a coarse-to-fine manner [18].

With those modalities in mind, we can now cover the process of segmenting the images.

2.2 Image Segmentation: From Images to Segmentation Maps

Image segmentation is the process of categorizing each pixel (or voxel) of an image as belonging to one of several predefined classes. For instance, a 3-class problem of segmenting CT images could be liver, liver tumour, and background. Each class gets assigned an arbitrary class label, such as '0', '1' and '2' for the background, liver and tumour, respectively. The segmentation output is another image of the same size, only the value for each voxel corresponds to the class label of that voxel. Voxels belonging to the liver would each have a value of '1', etcetera. This resulting image is called a **segmentation map**, since it acts as a map for the original image. In medical image segmentation commonly only target is segmented, which is called **binary segmentation** since we are using two classes, the target, and the background.

However, multi-class segmentation problems can be reduced to a set of binary segmentation problems where a separate class-vs-background segmentation map is constructed for each class. Mathematically, given a set of K classes, and an input d-dimensional image of N channels $I(A)$, $I \in \mathbb{R}^N$, $A \in \mathbb{Z}^d_{\geq 0}$ where A is a vector representing the location of each voxel, the segmentation map can be expressed as a function $M : \mathbb{Z}^d_{\geq 0} \to \mathbb{R}^K$ mapping each pixel location to a vector of class probabilities:

$$M(A) = (\ \Pr(C_0 \,|\, I(A)),\ \Pr(C_1 \,|\, I(A)),\ \cdots,\ \Pr(C_{K-1} \,|\, I(A))\ ) \tag{2.3}$$

Where $\Pr(C_i \,|\, I(A))$ denotes the probability that the voxel $I(A)$ contains an object of class C_i. Expressed this way, the segmentation map is a K-channel image of the same size as I, where each channel corresponds to a probability map of finding an object of a given class at a given voxel location.

In the case of binary segmentation where C_0 is the background and the C_1 is the target object, $Pr(C_1 \,|\, (A)) = 1 - Pr(C_0 \,|\, I(A))\ \forall A$ and therefore $M(A)$ can be treated as a scalar value $M(A) = Pr(C_1 \mid I(A))$. This representation is very common as many medical image segmentation problems are binary segmentation problems, such as segmenting organs, cell nuclei, skin lesions, etc.

Ultimately, to delineate tissues the segmentation masks $M(A)$ are commonly binarized such that voxels containing the target object become '1' and the background becomes '0'. This is why $M(A)$ is also commonly called a **segmentation mask**. In computer vision, the term mask refers to a binary image $M_{01}(A) \in \{0, 1\}$ that hides (masks) regions in another image producing a masked image $I_m(A) = I(A)M_{01}(A)$. In the context

of deep learning-based image segmentation, the terms segmentation map and segmentation mask are often used interchangeably.

3 Research Approach

- In diesem Kapitel beschreiben Sie, was Sie getan haben, um die Ziele und die Ergebnisse der Arbeit zu erreichen. Hierzu gehört auch wie, womit, warum (zu welchem Zweck) von wem sonst noch und in welcher Reihenfolge die beschriebenen Aktivitäten durchgeführt wurden.

3.1 Section 1

3.1.1 Subsection 1.1

3.1.2 Subsection 1.2

3.2 Section 2

3.2.1 Subsection 2.1

3.2.2 Subsection 2.2

3.3 Section 3

4 Results

- Hier finden sich alle Ergebnisse.
- Die Ergebnisse sind durch Messungen, Formeln, Diagramme, Tabellen und eine logische, für den Leser nachvollziehbare Argumentation zu belegen.
- Kein hierfür erforderlicher Beleg (etwa eine Tabelle oder ein Diagramm) wird in den Anhang "'verbannt"'.
- Die Kernergebnisse müssen in Abbildungen, Tabellen und Formeln stecken.
- Der Leser muss jedes Ergebnis anhand der gegebenen Zahlen, Daten und Fakten (ZDF) einfach nachvollziehen können, ohne langwierige eigenen Rechnungen.
- Im diesem Kapitel dürfen keine grundlegenden Erläuterungen und Begründungen mehr auftauchen.
- Wenn alle Ergebnisse beschrieben sind, steht fest, welche Grundlagen benötigt werden und wie das Vorgehen hierfür zu beschreiben ist, damit der Leser es mit Blick auf die Ergebnisse nachvollziehen kann.

4.1 Section 1

4.1.1 Subsection 1.1

4.1.2 Subsection 1.2

4.2 Section 2

4.2.1 Subsection 2.1

4.2.2 Subsection 2.2

4.3 Section 3

5 Discussion

6 Conclusion

- Hier wird die gesamte Arbeit auf einer Seite, maximal zwei Seiten zusammengefasst.
- Insbesondere werden die zentralen Ergebnisse dargestellt.
- Es werden Vorschläge beschrieben zur Lösung der offen gebliebenen Fragen, oder derer, die sich im Laufe der Arbeit ergeben haben.
- In diesem Kapitel folgt auch eine kritische Analyse der eigenen Arbeit.
- Ein Ausblick auf die Bearbeitung offener Fragen kann gegeben werden.

Bibliography

[1] D. Selle, B. Preim, A. Schenk, and H.-O. Peitgen, "Analysis of vasculature for liver surgical planning", *IEEE Transactions on Medical Imaging*, vol. 21, no. 11, pp. 1344–1357, Nov. 2002, ISSN: 0278-0062. DOI: 10.1109/TMI.2002.801166. (visited on 09/27/2023).

[2] S. Devunooru, A. Alsadoon, P. W. C. Chandana, and A. Beg, "Deep learning neural networks for medical image segmentation of brain tumours for diagnosis: A recent review and taxonomy", *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 1, pp. 455–483, Jan. 2021, ISSN: 1868-5137, 1868-5145. DOI: 10.1007/s12652-020-01998-w. (visited on 09/27/2023).

[3] Z. Sobhaninia, S. Rafiei, A. Emami, N. Karimi, K. Najarian, S. Samavi, and S. M. Reza Soroushmehr, "Fetal Ultrasound Image Segmentation for Measuring Biometric Parameters Using Multi-Task Deep Learning", in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany: IEEE, Jul. 2019, pp. 6545–6548, ISBN: 978-1-5386-1311-5. DOI: 10.1109/EMBC.2019.8856981. (visited on 09/27/2023).

[4] G. Bastarrika, J. Broncano, U. J. Schoepf, F. Schwarz, Y. S. Lee, J. A. Abro, P. Costello, and P. L. Zwerner, "Relationship Between Coronary Artery Disease and Epicardial Adipose Tissue Quantification at Cardiac CT", *Academic Radiology*, vol. 17, no. 6, pp. 727–734, Jun. 2010, ISSN: 10766332. DOI: 10.1016/j.acra.2010.01.015. (visited on 09/27/2023).

[5] M. Antonelli, A. Reinke, S. Bakas, K. Farahani, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, O. Ronneberger, R. M. Summers, B. Van Ginneken, M. Bilello, P. Bilic, P. F. Christ, R. K. G. Do, M. J. Gollub, S. H. Heckers, H. Huisman, W. R. Jarnagin, M. K. McHugo, S. Napel, J. S. G. Pernicka, K. Rhode, C. Tobon-Gomez, E. Vorontsov, J. A. Meakin, S. Ourselin, M. Wiesenfarth, P. Arbeláez, B. Bae, S. Chen, L. Daza, J. Feng, B. He, F. Isensee, Y. Ji, F. Jia, I. Kim, K. Maier-Hein, D. Merhof, A. Pai, B. Park, M. Perslev, R. Rezaiifar, O. Rippel, I. Sarasua, W. Shen, J. Son, C. Wachinger, L. Wang, Y. Wang, Y. Xia, D. Xu, Z. Xu, Y. Zheng, A. L. Simpson, L. Maier-Hein, and M. J. Cardoso, "The Medical Segmentation Decathlon", *Nature Communications*, vol. 13, no. 1, p. 4128, Jul. 2022, ISSN: 2041-1723. DOI: 10.1038/s41467-022-30695-9. (visited on 09/28/2023).

[6] C. Edlund, T. R. Jackson, N. Khalid, N. Bevan, T. Dale, A. Dengel, S. Ahmed, J. Trygg, and R. Sjögren, "LIVECell—A large-scale dataset for label-free live cell segmentation", *Nature Methods*, vol. 18, no. 9, pp. 1038–1045, Sep. 2021, ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-021-01249-6. (visited on 09/28/2023).

[7] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler, and A. Halpern, *Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)*, Mar. 2019. arXiv: 1902.03368 [cs]. (visited on 09/28/2023).

[8] J. Cho, K. Lee, E. Shin, G. Choy, and S. Do, *How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?*, Jan. 2016. arXiv: 1511.06348 [cs]. (visited on 09/28/2023).

[9] E. Lotan, C. Tschider, D. K. Sodickson, A. L. Caplan, M. Bruno, B. Zhang, and Y. W. Lui, "Medical Imaging and Privacy in the Era of Artificial Intelligence: Myth, Fallacy, and the Future", *Journal of the American College of Radiology*, vol. 17, no. 9, pp. 1159–1162, Sep. 2020, ISSN: 15461440. DOI: 10.1016/j.jacr.2020.04.007. (visited on 09/28/2023).

[10] B. Lutnick, B. Ginley, D. Govind, S. D. McGarry, P. S. LaViolette, R. Yacoub, S. Jain, J. E. Tomaszewski, K.-Y. Jen, and P. Sarder, "An integrated iterative annotation technique for easing neural network training in medical image analysis", *Nature Machine Intelligence*, vol. 1, no. 2, pp. 112–119, Feb. 2019, ISSN: 2522-5839. DOI: 10.1038/s42256-019-0018-3. (visited on 09/28/2023).

[11] R. Fosbinder and D. Orth, *Essentials of Radiologic Science*. Wolters Kluwer Health/Lippincott Williams & Wilkins, 2011, ISBN: 978-0-7817-7554-0.

[12] S. Kamalian, M. H. Lev, and R. Gupta, "Computed tomography imaging and angiography – principles", in *Handbook of Clinical Neurology*, vol. 135, Elsevier, 2016, pp. 3–20, ISBN: 978-0-444-53485-9. DOI: 10.1016/B978-0-444-53485-9.00001-5. (visited on 09/29/2023).

[13] A. A. Mahabadi, B. Balcer, I. Dykun, M. Forsting, T. Schlosser, G. Heusch, and T. Rassaf, "Cardiac computed tomography-derived epicardial fat volume and attenuation independently distinguish patients with and without myocardial infarction", *PLOS ONE*, vol. 12, no. 8, M. W. Merx, Ed., e0183514, Aug. 2017, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0183514. (visited on 09/29/2023).

[14] M. Benčević, I. Galić, M. Habijan, and A. Pižurica, "Recent Progress in Epicardial and Pericardial Adipose Tissue Segmentation and Quantification Based on Deep Learning: A Systematic Review", *Applied Sciences*, vol. 12, no. 10, p. 5217, May 2022, ISSN: 2076-3417. DOI: 10.3390/app12105217. (visited on 09/29/2023).

[15] H. Q. Nguyen, K. Lam, L. T. Le, H. H. Pham, D. Q. Tran, D. B. Nguyen, D. D. Le, C. M. Pham, H. T. T. Tong, D. H. Dinh, C. D. Do, L. T. Doan, C. N. Nguyen, B. T. Nguyen, Q. V. Nguyen, A. D. Hoang, H. N. Phan, A. T. Nguyen, P. H. Ho, D. T. Ngo, N. T. Nguyen, N. T. Nguyen, M. Dao, and V. Vu, *VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations*, 2020. arXiv: 2012.15029 [eess.IV].

[16] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, J. Seekins, D. A. Mong, S. S. Halabi, J. K. Sandberg, R. Jones, D. B. Larson, C. P. Langlotz, B. N. Patel, M. P. Lungren, and A. Y. Ng, *CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison*, Jan. 2019. arXiv: 1901.07031 [cs, eess]. (visited on 10/02/2023).

[17] V. Rotemberg, N. Kurtansky, B. Betz-Stablein, L. Caffery, E. Chousakos, N. Codella, M. Combalia, S. Dusza, P. Guitera, D. Gutman, A. Halpern, B. Helba, H. Kittler, K. Kose, S. Langer, K. Lioprys, J. Malvehy, S. Musthaq, J. Nanda, O. Reiter, G. Shih, A. Stratigos, P. Tschandl, J. Weber, and H. P. Soyer, "A patient-centric dataset of images and metadata for identifying melanomas using clinical context", *Scientific Data*, vol. 8, no. 1, p. 34, Jan. 2021, ISSN: 2052-4463. DOI: 10.1038/s41597-021-00815-z. (visited on 10/02/2023).

[18] A. Jha, H. Yang, R. Deng, M. E. Kapp, A. B. Fogo, and Y. Huo, "Instance segmentation for whole slide imaging: End-to-end or detect-then-segment", *Journal of Medical Imaging*, vol. 8, no. 01, Jan. 2021, ISSN: 2329-4302. DOI: 10.1117/1.JMI.8.1.014001. (visited on 10/02/2023).