

Mass Segmentation in Automated 3D Breast Ultrasound using Deep Learning

H. Fayyaz¹, M. Soryani^{2*}, E. Koozegar³, T. Tan⁴

¹M.Sc. Student, Artificial Intelligence Department, School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran

²Associate Professor, Artificial Intelligence Department, School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran

³Ph.D. Student, Artificial Intelligence Department, School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran

⁴Postdoctoral Researcher, Department of Radiology and Nuclear Medicine, Radboud University, Nijmegen, Netherlands

Receipt in the Online Submission System: 28 July 2018, Received in Revised Form: 2 September 2018, Accepted: 27 September 2018

Abstract

Automated 3D breast ultrasound (ABUS) is a novel system for breast screening. It has been proposed as a supplementary modality to mammography for detection and diagnosis of breast cancers. Although ABUS has better performance for dense breasts, reading ABUS images is time-consuming and exhausting. A computer-aided detection (CAD) system can be helpful for interpretation of ABUS images. Mass Segmentation in CADE and CADx systems play the leading role because it affects the performance of succeeding stages. Besides, it is a very challenging task because of the vast variety in size, shape, and texture of masses. Moreover, imbalanced datasets make segmentation harder. A novel mass segmentation approach based on deep learning is introduced in this paper. The deep network that is used in this study for image segmentation is inspired by U-net which has been used broadly for dense segmentation in recent years. Performance was determined using a dataset of 50 masses including 38 malignant and 12 benign masses.

Keywords: *Breast Cancer, Automated Diagnosis, Tumor, 3D ABUS, Mass Segmentation, Deep Learning*

*Corresponding Author

Address: Artificial Intelligence Department, School of Computer Engineering, Iran University of Science and Technology, P. O. Box: 1684613114, Tehran, Iran

Tel: +98-21-73225304

Fax: +98-21-73021220

E-mail: soryani@iust.ac.ir

قطعه‌بندی توده‌ها در تصاویر سه‌بعدی ABUS به کمک یادگیری ژرف

حامد فیاض^۱، محسن سریانی^{۲*}، احسان‌اله کوزه‌گر^۳، تائو تن^۴

^۱ دانشجوی کارشناسی ارشد، گروه هوش مصنوعی، دانشکده‌ی مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران

^۲ دانشیار، گروه هوش مصنوعی، دانشکده‌ی مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران

^۳ دانشجوی دکتری، گروه هوش مصنوعی، دانشکده‌ی مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران

^۴ پژوهش‌گر پسادکتری، دانشکده‌ی رادیولوژی و پزشکی هسته‌ای، مرکز پزشکی دانشگاه رادبود، نایمخن، هلند

تاریخ ثبت در سامانه: ۱۳۹۷/۵/۶، بازنگری: ۱۳۹۷/۶/۱۱، پذیرش قطعی: ۱۳۹۷/۷/۵

چکیده

سرطان پستان دومین عامل مرگ زنان در جهان است. هر چه این بیماری زودتر تشخیص داده شود، احتمال موفقیت در درمان آن بیش‌تر خواهد بود. امروزه به دلیل مشکلات تصویربرداری ماموگرافی، استفاده از تصاویر فراصوت برای تشخیص سرطان پستان در حال افزایش است. یکی از انواع سامانه‌های تصویربرداری پستان، ABUS است. این نوع تصویربرداری از مزایای زیادی نسبت به ماموگرافی و سایر روش‌های تصویربرداری فراصوت برخوردار می‌باشد. طراحی یک سامانه‌ی تشخیص به کمک کامپیوتر برای تحلیل تصاویر فراصوت در کنار رایج شدن این نوع تصویربرداری ضروری است. این سامانه‌ها معمولاً دارای چهار بخش پیش‌پردازش، قطعه‌بندی، استخراج ویژگی و دسته‌بندی هستند. افزایش دقت قطعه‌بندی، باعث افزایش دقت عمل‌کرد سامانه می‌شود. در پژوهش‌های پیشین، از روش‌های پویس حلزونی و کانتور فعال برای قطعه‌بندی توده‌ها در تصاویر سه‌بعدی استفاده شده است. در سال‌های اخیر، استفاده از یادگیری ژرف در زمینه‌های مختلف، منجر به کسب نتایج قابل توجهی شده است که امکان دستیابی به این نتایج با روش‌های سنتی پیشین وجود نداشت. در این پژوهش، با استفاده از یک شبکه‌ی عصبی ژرف، که از معماری 3D U-net بهره می‌برد، توده‌های سرطانی در تصاویر ABUS قطعه‌بندی شده‌اند. در این روش، به منظور بهبود عمل‌کرد شبکه، از یک رویکرد جدید برای پس‌پردازش استفاده شده است. مجموعه‌ی دادگان مورد استفاده، از ۳۲ بیمار جمع‌آوری شده و شامل ۵۰ توده (۳۸ توده‌ی بدخیم و ۱۲ توده‌ی خوش‌خیم) است. برای ارزیابی دقت قطعه‌بندی، از معیار ضریب Dice استفاده شده است. میانگین دقت به دست آمده روی مجموعه‌ی دادگان مورد استفاده در این پژوهش، ۰/۷۷ است.

کلیدواژه‌ها: سرطان پستان، توده، تصاویر سه‌بعدی فراصوت، قطعه‌بندی، یادگیری ژرف

*نویسنده مسئول

نشانی: گروه هوش مصنوعی، دانشکده‌ی مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران، کد پستی: ۱۶۸۴۶۱۳۱۱۴

تلفن: +۹۸ (۲۱) ۷۳۲۲۵۳۰۴

دورنگار: +۹۸ (۲۱) ۷۳۰۲۱۲۲۰

پست الکترونیکی: soryani@iust.ac.ir

۱- مقدمه

شایع‌ترین سرطان در بین زنان، سرطان پستان است. هم‌چنین علت اصلی مرگ‌ومیرهای سرطانی در بین زنان نیز سرطان پستان می‌باشد [۱]. اگر این سرطان سریع و به موقع تشخیص داده شود، مرگ‌ومیر ناشی از آن کاهش یافته و هم‌چنین بیماران از جهت عمل جراحی و شیمی‌درمانی، دشواری‌های کم‌تری را تحمل می‌کنند. یکی از روش‌های تشخیص سرطان پستان، تصویربرداری پزشکی بوده که در این میان، ماموگرافی رایج‌ترین روش تصویربرداری از پستان است. امروزه روش‌های تصویربرداری زیادی در حال توسعه هستند که می‌توان از آن‌ها برای تشخیص زودهنگام این بیماری استفاده کرد. یکی از مهم‌ترین این روش‌ها، ABUS^۱ است، که نسبت به ماموگرافی، دارای مزایایی بوده که در این‌جا به چند نمونه از آن‌ها اشاره شده است:

(۱) امواج فراصوت مورد استفاده در ABUS، ضرر بسیار کم‌تری نسبت به اشعه‌ی ایکس مورد استفاده در ماموگرافی دارند
(۲) حساسیت ماموگرافی روی زنانی که پستان‌های متراکم دارند کم بوده و در این موارد استفاده از سامانه‌های ABUS منجر به ایجاد حساسیت بیش‌تری می‌شود
(۳) ماموگرافی نسبت به ABUS منجر به پاسخ‌های مثبت نادرست بیش‌تری می‌شود

با توجه به مزیت‌های ABUS نسبت به ماموگرافی و سایر روش‌های تصویربرداری از پستان، نیاز به توسعه‌ی یک سامانه‌ی تشخیص به کمک کامپیوتر، برای این نوع تصاویر ضروری به نظر می‌رسد.

سامانه‌ی تشخیص به کمک کامپیوتر، دارای چهار بخش پیش‌پردازش، قطعه‌بندی، استخراج ویژگی‌ها و دسته‌بندی است. در بخش قطعه‌بندی، با توجه به محل تقریبی توده، که توسط کاربر و یا ماشین تعیین شده است، قطعه‌بندی توده انجام شده تا مرزهای دقیق آن مشخص گردد و در نتیجه، در مرحله‌ی بعد، امکان استخراج ویژگی‌های دقیقی از آن فراهم شود. اهمیت قطعه‌بندی به این دلیل است که هرچه قطعه‌بندی دقیق‌تر باشد، سامانه در مراحل بعد عمل‌کرد بهتری خواهد داشت. در این پژوهش، روشی بر مبنای یادگیری ژرف، جهت قطعه‌بندی توده‌ها در تصاویر سه‌بعدی ABUS، ارائه شده است.

۲- کارهای پیشین

به طور کلی، از روش‌های زیر برای قطعه‌بندی تصاویر فراصوت پستان استفاده می‌شود [۲]:

- روش‌های مبتنی بر آستانه‌گذاری
- روش‌های مبتنی بر خوشه‌بندی
- روش‌های مبتنی بر آب‌پخشان
- روش‌های مبتنی بر گراف
- روش‌های مبتنی بر کانتور فعال
- روش‌های مبتنی بر شبکه‌های عصبی

در این بخش، ابتدا به معرفی روش‌های پایه‌ی ارائه شده برای قطعه‌بندی تصاویر سه‌بعدی ABUS پرداخته شده، سپس روش‌های مبتنی بر یادگیری ژرف برای قطعه‌بندی تصاویر پزشکی ارایه شده است.

۲-۱- قطعه‌بندی تصاویر ABUS با روش‌های پایه

کانتور فعال، زیربنای دو روش ارائه شده برای قطعه‌بندی تصاویر ABUS است. از این روش برای مشخص کردن مرزهای یک جسم در تصویر استفاده می‌شود. در این روش، یک تابع انرژی برای کانتورها تعریف کرده و سعی می‌شود تا آن تابع انرژی، کمینه گردد. این تابع انرژی از مجموع دو انرژی درونی و انرژی بیرونی تشکیل شده است. انرژی درونی^۲، مربوط به ویژگی‌های کانتور، مانند منظم بودن شکل آن بوده و انرژی بیرونی^۳، مربوط به ویژگی‌های تصویر، مانند لبه‌ها، می‌باشد.

کو و هم‌کارانش در سال ۲۰۱۳ [۳]، از این روش برای قطعه‌بندی تصاویر ABUS استفاده کرده و به ضریب Dice ۰/۷۰ دست پیدا کردند.

کوزه‌گر و هم‌کارانش [۴]، با استفاده از یک مدل هندسی، موسوم به «تکامل مجموعه‌ی تراز با فاصله‌ی تنظیم شده»^۴، توده‌ها را در تصاویر سه‌بعدی ABUS، قطعه‌بندی نمودند.

آن‌ها در کار بعدی خود [۵]، روشی سه مرحله‌ای (شامل نویز زدایی، پیش‌قطعه‌بندی با روش رشد ناحیه‌ای و قطعه‌بندی دقیق با مدل‌های شکل‌پذیر سه‌بعدی) را برای قطعه‌بندی توده‌ها در تصاویر سه‌بعدی ABUS ارائه کردند.

تمامی تصاویر فراصوت، دارای نوعی نویز نقطه‌ای به نام speckle بوده، که این نویز باعث ایجاد مشکل در قطعه‌بندی

^۲ E_{Internal}^۳ E_{External}^۴ Distance Regularized Level Set Evolution^۱ Automated Breast UltraSound

- (۳) اضافه کردن چندین جهت اسکن در مدل حلزونی
 (۴) بازسازی سه‌بعدی
 (۵) افزودن اطلاعات عمق، جهت بهبود نتیجه‌ی قطعه‌بندی

۲-۲- قطعه‌بندی به کمک یادگیری ژرف

در سال‌های اخیر، از یادگیری ژرف در بسیاری از زمینه‌های هوش مصنوعی، از جمله پردازش تصویر، بهره گرفته می‌شود. به دلیل تنوع موارد کاربرد یادگیری ژرف، یک‌سان نبودن نیازها و پیچیدگی‌های موجود در زمینه‌های مختلف، در سال‌های اخیر معماری‌های بسیار متنوعی ارائه شده است. یادگیری ژرف در زمینه‌ی تحلیل اطلاعات در تصاویر پزشکی نیز کاربردهای زیادی دارد، که از آن جمله می‌توان به دسته‌بندی^۵، تشخیص^۶ و قطعه‌بندی اشاره نمود. در ادامه، به بررسی مهم‌ترین معماری‌های ارائه شده برای قطعه‌بندی تصاویر پزشکی پرداخته می‌شود. شایان ذکر است که تاکنون از هیچ معماری ژرفی برای قطعه‌بندی تصاویر سه‌بعدی ABUS استفاده نشده است.

در سال ۲۰۱۲، سیرسن^۷ و هم‌کارانش [۸] برای نخستین بار از شبکه‌ی عصبی ژرف برای قطعه‌بندی ساختار نورون‌ها در تصاویر میکروسکوپ الکترونی استفاده کردند. در این روش، از یک شبکه‌ی عصبی ژرف، به عنوان دسته‌بند پیکسلی، استفاده شده است. احتمال غشا بودن یا غشا نبودن برای پیکسل‌های این تصاویر، در نظر گرفته می‌شود. این شبکه، احتمال غشا بودن یک پیکسل را با استفاده از شدت روشنایی پیکسل‌ها در یک پنجره‌ی مربعی به مرکز پیکسل مذکور، تشخیص می‌دهد. از آن‌جا که هر تصویر، به صورت مجموعه‌ای از پیکسل‌ها می‌باشد، برای قطعه‌بندی باید تمام پیکسل‌های آن را به عنوان ورودی به شبکه داده تا نقشه‌ی کامل احتمال غشا بودن هر یک از پیکسل‌ها، به دست آید. ابتدا دسته‌بند با مجموعه‌ی تصاویر آموزش، آموزش داده می‌شود. سپس برای قطعه‌بندی تصویر تست، برای تک‌تک پیکسل‌ها دسته‌بندی انجام شده تا نقشه‌ی احتمال غشا بودن پیکسل‌ها، تولید شود. در ادامه، بعد از پس‌پردازش و اعمال مقدار آستانه، قطعه‌بندی دودویی به دست می‌آید.

در سال ۲۰۱۵، رونبرگر^۸ و هم‌کارانش مدل U-net را معرفی کردند [۹]. بر خلاف مدل قبل، U-net در هر مرحله از اجرای

تصاویر می‌شود. در [۵]، از روش OBNLM^۱ برای حذف نویز speckle استفاده شده است.

در مرحله‌ی پیش‌قطعه‌بندی، با آگاهی از دایروی بودن و حجم نمونه‌های آموزش، یک GMM^۲ ساخته می‌شود. از GMM به عنوان تابع چگالی احتمال، برای تعیین افزوده شدن یا افزوده نشدن وکسل‌های جدید به توده‌ی در حال رشد، استفاده می‌شود. بر اساس این الگوریتم، آستانه‌ای که باعث می‌شود شکل ناحیه دارای بیش‌ترین احتمال شود، به عنوان آستانه‌ی شباهت نهایی، در نظر گرفته می‌شود. در مرحله‌ی قطعه‌بندی، از مدل‌های شکل‌پذیر با تابع انرژی مطابق رابطه‌ی (۱) استفاده شده است [۵]:

$$E = E_{DRLSE} + E_{shape} \quad (1)$$

E_{DRLSE} ، به صورت رابطه‌ی (۲) بیان می‌شود [۵]:

$$E_{DRLSE} = \mu R_p(\Phi) + E_{ext}(\Phi) \quad (2)$$

که $R_p(\Phi)$ تنظیم‌کننده‌ی فاصله، μ ثابت معادله و $E_{ext}(\Phi)$ انرژی بیرونی است. هم‌چنین، E_{shape} نیز مطابق رابطه‌ی (۳) می‌باشد [۵]:

$$E_{shape} = \frac{1}{2} \int_{\Omega} (H_{\epsilon}(\Phi) - H_{\epsilon}(\Phi^*))^2 dx \quad (3)$$

معیار DSC^۴ به دست آمده برای این پژوهش، برابر ۰/۷۴ است. این ضریب، یک معیار شباهت بوده که برای ارزیابی دقت سامانه، مورد استفاده قرار می‌گیرد و فرمول آن در بخش نتایج ارائه خواهد شد.

در پژوهش سوم، تائو و هم‌کارانش، از یک روش برنامه‌ریزی پویا برای به دست آوردن مرز توده استفاده کرده و به ضریب Dice ۰/۷۳ دست یافتند [۶]. این روش با عنوان پویا حلزونی شناخته می‌شود که برای نخستین بار توسط ونگ و هم‌کارانش، به منظور بخش‌بندی ندول‌های ریوی در تصاویر سه‌بعدی سی‌تی‌اسکن معرفی گردید [۷]. به دلیل جزئیات زیاد این روش، در این‌جا تنها مراحل آن به اختصار بیان می‌شود. این الگوریتم شامل ۵ مرحله است:

- (۱) تبدیل حجم مورد نظر به یک تصویر دویعدی
 (۲) لبه‌یابی با استفاده از برنامه‌ریزی پویا

^۵ Classification

^۶ Diagnosis

^۷ Ciresan

^۸ Ronneberger

^۱ Optimized Bayesian Non-Local Means

^۲ Gaussian Mixture Model

^۳ Distance Regularized Level Set Evolution

^۴ Dice Similarity Coefficient

در ادامه، یک max-pooling $2 \times 2 \times 2$ قرار دارد. در مسیر expansive، در هر واحد از یک لایه‌ی upconvolution $2 \times 2 \times 2$ و دو کانولوشن که همراه ReLU هستند استفاده شده است. یکی از تفاوت‌های این نسخه با معماری اصلی، استفاده از Batch normalization، قبل از هر ReLU می‌باشد.

۳- مجموعه‌ی دادگان و روش پیشنهادی

۳-۱- مجموعه‌ی دادگان

تصاویر سه‌بعدی ABUS که در این پژوهش مورد استفاده قرار گرفته، توسط دو دستگاه Somov^۴ و ACUSON^۵ تهیه شده‌اند. در این مجموعه‌ی دادگان، تصاویر تهیه شده توسط دستگاه اول، حداکثر دارای ابعاد $14/6 \times 16/8$ سانتی‌متر در صفحه‌ی کروئال و عمق $4/86$ سانتی‌متر بوده و تصاویر تهیه شده توسط دستگاه دوم، حداکثر دارای ابعاد $15/4 \times 16/8$ سانتی‌متر در صفحه‌ی کروئال و عمق 6 سانتی‌متر می‌باشند. مبدل دستگاه اول، دارای فرکانس ثابت 8 یا 10 مگاهرتز بوده، اما مبدل دستگاه دوم، دارای فرکانسی متغیر بین 5 تا 14 مگاهرتز است. مجموعه‌ی دادگان موجود، دارای 42 حجم تصویری بوده که 35 حجم توسط دستگاه اول و 7 حجم توسط دستگاه دوم تهیه شده است. این تصاویر از 32 بیمار جمع‌آوری شده و شامل 50 توده (38 توده‌ی بدخیم و 12 توده‌ی خوش‌خیم) است. یکی از چالش‌ها در این نوع تصاویر، وجود تنوع در اندازه‌ی توده‌ها است. در این مجموعه‌ی دادگان نیز توده‌ها از نظر اندازه بسیار متنوع هستند که این مساله در شکل‌های (۱)، (۲) و (۳) قابل مشاهده می‌باشد. این نمودارها به ترتیب مربوط به نمایش تجمعی طول توده‌ها در راستای محور x ، y و z تصاویر می‌باشند.

۳-۲- روش پیشنهادی اول: شبکه‌ی 3D U-net

در روش اول، برای قطعه‌بندی توده‌ها در تصاویر ABUS، از یک شبکه‌ی 3D U-net استفاده می‌شود که در شکل (۴) قابل مشاهده است. همان‌طور که اشاره شد، این شبکه دارای یک مسیر contracting و یک مسیر expansive می‌باشد. هم‌چنین در این معماری از سه ارتباط پرشی^۶ برای استخراج بهتر ویژگی‌ها استفاده شده است.

شبکه، به جای یک پیکسل، مجموعه‌ای از پیکسل‌ها را بر حسب می‌زند. از نظر طراحان این شبکه، راه حل ارائه شده در [۸]، دارای دو ایراد است:

(۱) از آن‌جا که شبکه باید به ازای تک‌تک پیکسل‌ها اجرا شود، سرعت به دست آوردن خروجی کامل، بسیار پایین است. هم‌چنین، در patch-هایی که برای ارزیابی پیکسل‌های مجاور، به شبکه داده می‌شوند، افزونگی زیادی مشاهده می‌شود.

(۲) باید مصالحه‌ای بین استفاده از اطلاعات زمینه و استفاده از اطلاعات محلی انجام گیرد. هرچه اندازه‌ی patch بزرگ‌تر باشد، به لایه‌های max pooling بیش‌تری نیاز خواهد بود. از طرفی، انتخاب patch-های کوچک، باعث می‌شود که شبکه اطلاعات زمینه‌ای اندکی را مشاهده کند.

رونبرگر و هم‌کارانش برای رفع مشکلات ذکر شده، معماری U-net را پیشنهاد دادند.

سیسک^۱ و هم‌کارانش در سال ۲۰۱۶، شبکه‌ی 3D U-net، که نسخه‌ی گسترش یافته‌ی معماری U-net است، را ارائه کردند [۱۰]. از آن‌جا که در صفحه‌ی نمایش رایانه، تنها تصاویر دوبعدی قابل مشاهده است، برچسب گذاری و کسل‌ها و قطعه‌بندی تصاویر سه‌بعدی توسط رادیولوژیست قابل انجام نمی‌باشد. در نتیجه، این کار باید برش به برش انجام گیرد، که از کیفیت لازم برخوردار نخواهد بود. هم‌چنین، به دلیل زیاد بودن تعداد برش‌ها و تشابه برش‌های مجاور، انجام این کار وقت‌گیر و خسته‌کننده است. در نتیجه، برچسب گذاری کامل حجم‌های سه‌بعدی، راه‌کار مناسبی برای ساخت مجموعه‌ی دادگان بزرگ و غنی نخواهد بود. از این‌رو، شبکه‌ی 3D U-net به نحوی طراحی شده است که با دو نوع برچسب گذاری غیرکامل، قابلیت یادگیری دارد:

(۱) یک مجموعه‌ی دادگان با برچسب‌های پراکنده^۲ موجود باشد و شبکه‌ی برچسب گذاری آن را کامل کند
(۲) برای هر حجم موجود، چند برش برچسب خورده وجود داشته باشد و شبکه با یادگیری از آن‌ها، نمونه‌های جدید را به صورت کامل برچسب بزند.

این معماری مانند معماری U-net، دارای یک مسیر contracting و یک مسیر expansive است. در مسیر contracting، هر واحد شامل دو کانولوشن $3 \times 3 \times 3$ بوده که پس از آن از ReLU^۳ به عنوان تابع فعال‌سازی استفاده شده و

^۴ SonoV Automated 3D Breast Ultrasound System

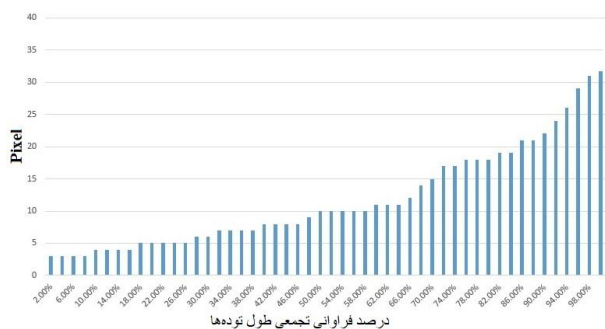
^۵ ACUSON S2000 Automated Breast Volume Scanning System

^۶ Concat

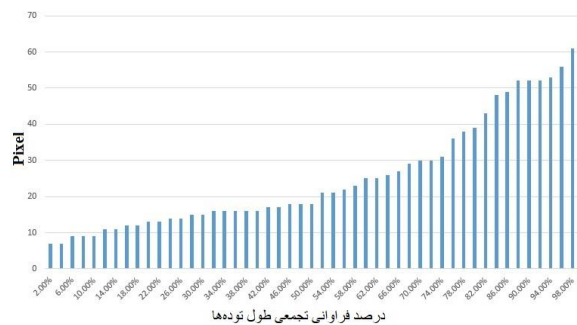
^۱ Cicek

^۲ Sparse

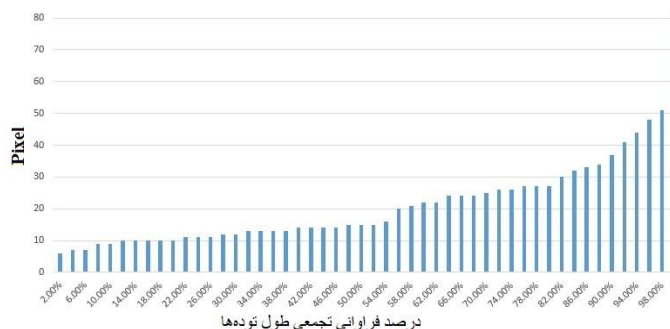
^۳ Rectified Linear Unit



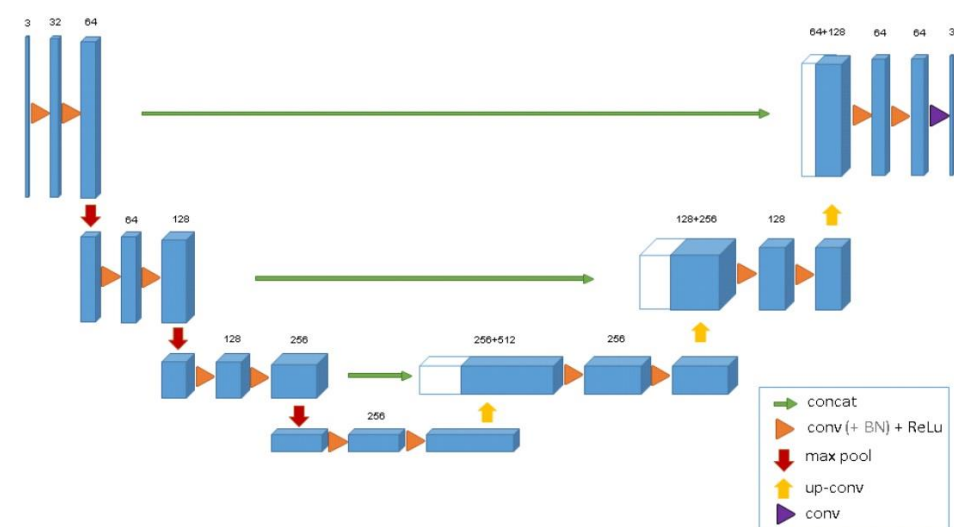
شکل (۲) - نمودار تجمعی طول توده‌ها در محور Y تصاویر



شکل (۱) - نمودار تجمعی طول توده‌ها در محور X تصاویر



شکل (۳) - نمودار تجمعی طول توده‌ها در محور Z تصاویر



شکل (۴) - معماری 3D U-net [۴]

مثبت وجود دارد، که این نسبت بسیار نامتوازن بوده و به نظر می‌رسد که چنین مجموعه‌ی دادگانی برای آموزش یک شبکه‌ی عصبی مناسب نباشد.

یکی از روش‌هایی که موجب متوازن شدن مجموعه‌ی دادگان می‌شود، انتخاب یک پنجره‌ی محدود در اطراف توده، به جای انتخاب کل تصویر سه‌بعدی، جهت ارائه به شبکه‌ی عصبی می‌باشد. حداکثر اندازه‌ی توده‌ها در حجم‌های موجود، $80 \times 80 \times 32$ پیکسل است. بنابراین، برای سهولت کار می‌توان به جای استفاده از کل تصویر در ورودی شبکه‌ی عصبی، از پنجره‌های $80 \times 80 \times 32$ ، که به مرکزیت توده بریده می‌شوند،

اندازه‌ی تقریبی حجم‌های موجود در مجموعه‌ی دادگان، $270 \times 270 \times 80$ پیکسل است. ساده‌ترین روش، وارد کردن کل حجم به ورودی شبکه است. به‌ازای هر ورودی، 5832000 وکسل وجود دارد که شبکه باید آن‌ها را به صورت هم‌زمان پردازش کند. دو مشکل اساسی در این مرحله وجود دارد:

(۱) به دلیل ابعاد بزرگ پنجره، آموزش چنین شبکه‌ای به منابع سخت‌افزاری زیادی احتیاج داشته و بسیار زمان‌بر خواهد بود (۲) با توجه به نامتوازن بودن تعداد وکسل‌های توده و غیرتوده، نسبت نمونه‌های مثبت به نمونه‌های منفی در حدود $0.01/0$ خواهد بود. یعنی به ازای هر 1000 وکسل منفی، یک وکسل

می‌توان پنجره را کوچک‌تر از $32 \times 80 \times 80$ نیز در نظر گرفت، اما هرچه اندازه‌ی پنجره کوچک‌تر شود، بخش کم‌تری از توده‌هایی که بزرگ‌تر از اندازه‌ی پنجره هستند، در پنجره جای گرفته و قطعه‌بندی می‌شوند و در نتیجه، دقت قطعه‌بندی کم‌تر خواهد شد. در روش دوم، رویکرد نوین و خلاقانه‌ای برای این مساله ارائه شده است که در ادامه توضیح داده خواهد شد. پیشنهاد می‌شود که اندازه‌ی پنجره، کوچک‌تر از بزرگ‌ترین توده ($32 \times 80 \times 80$) در نظر گرفته شود، به نحوی که نسبت وکسل‌های مثبت به وکسل‌های منفی، تا حد ممکن متوازن باشد. به عنوان مثال، اندازه‌ی پنجره برابر با $16 \times 32 \times 32$ در نظر گرفته می‌شود. در مرحله‌ی آموزش، مانند بخش قبل، تصاویر اصلی و افزونه‌های آن‌ها با اندازه‌ی پنجره‌ی $16 \times 32 \times 32$ ، به شبکه داده شده و از شبکه انتظار می‌رود تا تمامی وکسل‌ها را مطابق با یک درستی مرجع، برچسب بزند. در مرحله‌ی تست، با توجه به این‌که اندازه‌ی پنجره، کوچک‌تر از اندازه‌ی بزرگ‌ترین توده در نظر گرفته شده است، احتمال وقوع دو حالت وجود دارد. در حالت اول، اندازه‌ی توده‌ی مورد آزمایش، کوچک‌تر از اندازه‌ی پنجره بوده و کل توده شناسایی می‌شود. در حالت دوم، اندازه‌ی توده‌ی مورد آزمایش، بزرگ‌تر از پنجره بوده و در نتیجه، توده‌ی قطعه‌بندی شده بر یک یا چند وجه از پنجره‌ی سه‌بعدی مماس خواهد شد. در صورت وقوع این حالت، با کوچک‌نمایی مرحله‌به‌مرحله‌ی تصویر ورودی، توده را کوچک کرده تا در پنجره جای گیرد و سپس با بزرگ‌نمایی، خروجی شبکه به اندازه‌ی قبل از کوچک‌نمایی بازگردانده می‌شود. برای آگاهی از قرارگیری توده در پنجره‌ی انتخابی، می‌توان تعداد وکسل‌هایی که در دیواره‌ی خروجی شبکه، به عنوان توده برچسب خورده‌اند را محاسبه کرد. حال اگر تعداد آن‌ها از یک مقدار آستانه بیشتر باشد، مشخص می‌شود که توده‌ی ارزیابی شده در حجم مورد نظر قرار نگرفته و بنابراین یک مرحله‌ی کوچک‌نمایی دیگر روی حجم ورودی اعمال خواهد شد. الگوریتم پیشنهادی برای مرحله‌ی تست به صورت زیر است:

- ۱- حجمی که قرار است قطعه‌بندی شود، در یک پنجره‌ی سه‌بعدی به ورودی شبکه داده شود
- ۲- خروجی شبکه، به ازای ورودی داده شده، گرفته شود
- ۳- در خروجی شبکه، تعداد برچسب‌های مثبت در وکسل‌های دیواره‌ی حجم محاسبه شود
- ۴- اگر تعداد برچسب‌های مثبت در وکسل‌های دیواره‌ی حجم، بیش‌تر از مقدار آستانه بود، حجم مورد نظر ورودی با مقیاس

استفاده نمود. با این کار، سرعت آموزش شبکه افزایش یافته، میزان حافظه‌ی مورد نیاز برای آموزش شبکه، کاهش پیدا کرده و هم‌چنین نسبت نمونه‌های مثبت به نمونه‌های منفی، از $0/001$ به $0/028$ افزایش می‌یابد.

در مرحله‌ی آموزش، تصاویر اصلی و افزونه‌های^۱ آن‌ها به شبکه داده شده و از شبکه انتظار می‌رود تا تمامی وکسل‌ها را مانند یک درستی مرجع^۲، برچسب بزند. سپس در مرحله‌ی تست نیز حجم‌های $32 \times 80 \times 80$ به مرکزیت توده، بدون نمونه‌کاهی به شبکه داده شده و نتایج قطعه‌بندی آن‌ها، در خروجی دریافت می‌شود. فرض شده است که مرکز توده از قبل مشخص بوده و توسط یک متخصص و یا یک سامانه‌ی شناسایی توده، تعیین شده است.

به منظور یادگیری بهتر شبکه، از روش‌های زیر برای افزایش داده‌ها^۳ استفاده شده است. استفاده از این روش‌ها، نتایج متفاوتی به همراه داشته که در بخش ۴ به آن‌ها اشاره خواهد شد.

- نمونه‌افزایی^۴ با مقیاس‌های ۲ و ۴ و نمونه‌کاهی^۵ با مقیاس‌های $0/9$ ، $0/8$ ، $0/7$ ، $0/6$ ، $0/5$ و $0/4$
- چرخش تصویر در زوایای 45 ، 90 ، 135 ، 180 ، 225 ، 270 و 315
- وارونگی^۶

۳-۳- روش پیشنهادی دوم: پس پردازش

همان‌طور که اشاره شد، چالش اصلی در این کار، انتخاب اندازه‌ی پنجره است. هرچه اندازه‌ی پنجره بزرگ‌تر باشد، نسبت تعداد وکسل‌های منفی به وکسل‌های مثبت، بیش‌تر شده و مجموعه‌ی دادگان، نامتوازن‌تر می‌گردد. بنابراین، افزایش اندازه‌ی پنجره، موجب کاهش کیفیت یادگیری شبکه می‌شود. گرچه استفاده از پنجره‌ی $32 \times 80 \times 80$ در اطراف توده، به جای استفاده از کل تصویر سه‌بعدی، شرایط بهتری را برای آموزش شبکه فراهم ساخته است، اما به‌نظر می‌رسد که وجود 28 وکسل توده در کنار هر 1000 وکسل غیرتوده، شرایط مطلوبی را برای آموزش شبکه به وجود نیاورده است. برای افزایش نسبت وکسل‌های توده به وکسل‌های غیرتوده،

^۱ Augmented

^۲ Ground Truth

^۳ Data Augmentation

^۴ Up-Sampling

^۵ Down-Sampling

^۶ Mirroring

اکنون، ضریب Dice با توجه به پاسخ شبکه و یک درستی مرجع نمونه‌گاهی شده، محاسبه می‌شود.

۴- یافته‌ها و بحث

۴-۱- شرایط آموزش شبکه

برای آموزش شبکه، از بهینه‌ساز Adam [۱۱] استفاده شده است. این الگوریتم بهینه‌سازی، پارامترهای شبکه در هر اپک را با به دست آوردن میانگین هر دو مقدار شیب و مومنتوم دوم گرادین، به‌روز می‌کند. نرخ یادگیری برای هر یک از پارامترهای شبکه، حفظ شده و به طور جداگانه در طول آموزش تغییر پیدا می‌کند. مهم‌ترین مزیت Adam نسبت به AdaGrad و RMSProp، مقاوم بودن آن در برابر خلوت شدن گرادین‌های خطا، در زمان نزدیک شدن به پایان آموزش، می‌باشد.

۴-۲- معیار ارزیابی

به منظور ارزیابی نتایج، از روش 5-fold-cross-validation استفاده شده و برای ارزیابی دقت شبکه، از معیار DSC استفاده شده است (رابطه ۴).

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (4)$$

TP-ها، وکسل‌های توده هستند که به درستی تشخیص داده شده‌اند. FP-ها، وکسل‌های غیرتوده هستند که به اشتباه به عنوان توده تشخیص داده شده‌اند و FN-ها، وکسل‌های توده هستند که به اشتباه به عنوان غیرتوده تشخیص داده شده‌اند.

۴-۳- نتایج افزونگی داده‌ها

برای افزونگی داده‌ها، از چرخش، تغییر مقیاس و وارونگی استفاده شده است. نتایج آزمایش‌ها با استفاده از روش‌های افزونگی مختلف برای روش اول، در جدول (۱) ارائه شده است. همان‌طور که مشاهده می‌شود، افزونگی داده‌های حاصل از چرخش، نتیجه‌ی مناسبی نداشته و بنابراین، در ادامه از آن استفاده نشده است. ترکیب وارونگی و تغییر مقیاس با مقیاس‌های ۲ و ۵/۰، نتیجه‌ی بهتری نسبت به سایر حالت‌ها داشته است. از این‌رو، در سایر آزمایش‌ها نیز از این ترکیب به همراه تغییر مقیاس‌های بیشتر (با مقیاس‌های ۰/۹، ۰/۸، ۰/۷، ۰/۶ و ۰/۵)، استفاده شده است.

بعدی کوچک‌نمایی شده، به شبکه داده شده و مجدداً به مرحله‌ی ۲ بازگردانده شود.

۵- اگر تعداد برچسب‌های مثبت در وکسل‌های دیواره‌ی حجم، کم‌تر از مقدار آستانه بود، خروجی شبکه بزرگ‌نمایی شده و به عنوان پاسخ شبکه در نظر گرفته شود.

با این روش از بزرگ شدن پنجره و نامتوازن شدن بیش از حد دادگان، جلوگیری شده و هم‌چنین تمامی حجم‌های موجود در مجموعه‌ی دادگان، به صورت کامل قطعه‌بندی می‌شوند.

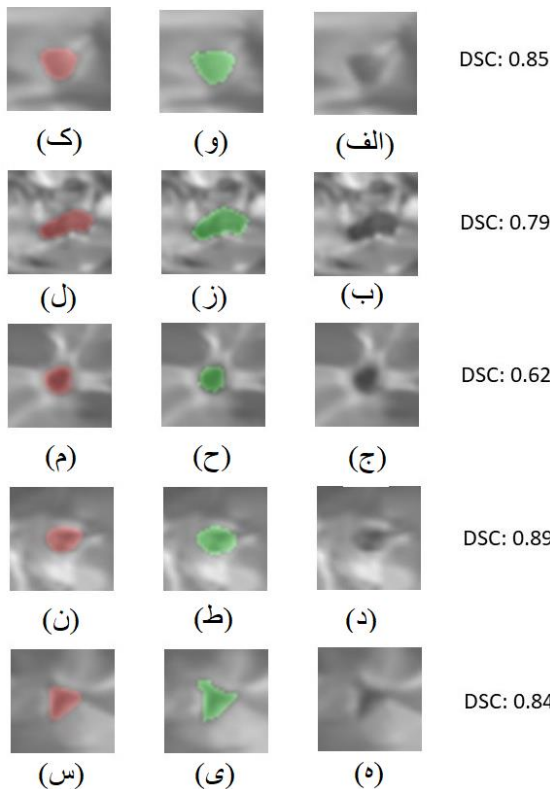
اکنون برای درک بهتر مطلب، مراحل آموزش و تست شبکه، با اندازه‌ی پنجره‌ی ۴۰×۴۰×۱۶، به تفصیل بیان خواهد شد.

(۱) ابتدا مجموعه‌ی دادگان به صورت تصادفی به ۵ قسمت مساوی تقسیم شده و در هر مرتبه، ۴ قسمت برای آموزش و ۱ قسمت برای تست در نظر گرفته می‌شود.

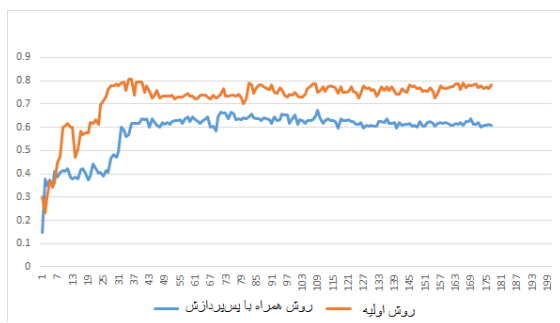
(۲) به ازای هر قسمت از دادگان که برای تست در نظر شده فرایند آموزش و تست، تکرار می‌شود. به منظور آموزش، ابتدا دادگان آموزش با روش‌های ارائه شده در بخش ۳-۲، افزون شده و سپس به ازای هر نمونه در مجموعه‌ی آموزش، یک پنجره با ابعاد ۴۰×۴۰×۱۶ به مرکزیت توده، برش داده می‌شود، که این پنجره‌ها به عنوان ورودی شبکه در نظر گرفته می‌شوند. در نهایت، شبکه با استفاده از دادگان آموزش و افزونگی‌های آن‌ها، آموزش داده می‌شود.

(۳) در مرحله‌ی تست، به ازای هر نمونه در مجموعه‌ی تست، یک پنجره با ابعاد ۴۰×۴۰×۱۶ به مرکزیت توده برش داده شده، بدون نمونه‌گاهی به شبکه داده شده و پاسخ آن دریافت می‌شود. اگر در دیواره‌ی حجم خروجی شبکه، هیچ وکسلی با برچسب مثبت وجود نداشته باشد، مقدار ضریب Dice با توجه به یک درستی مرجع محاسبه شده و اعمال مراحل برای نمونه‌ی بعدی شروع می‌شود. اما اگر در پاسخ شبکه، وکسلی با برچسب مثبت در دیواره‌ی حجم خروجی وجود داشت، نمونه‌گاهی با اولین مقیاس مشخص شده روی نمونه‌ی مورد تست اعمال می‌شود. سپس دوباره‌ی یک پنجره‌ی ۴۰×۴۰×۱۶ از نمونه‌ی تغییر مقیاس داده شده به مرکزیت توده، برش داده شده، پنجره‌ی برش داده شده‌ی جدید، به ورودی شبکه داده شده و وجود وکسل‌های مثبت در دیواره، مجدداً مورد بررسی قرار می‌گیرد. در صورت وجود وکسل مثبت در دیواره‌ی حجم خروجی شبکه، نمونه‌گاهی با مقیاس بعدی انجام می‌شود. انجام این کار تا دست‌یابی به یک ورودی از شبکه که خروجی آن وکسل مثبتی در دیواره نداشته باشد، ادامه داده می‌شود.

3D U-net، که پیش‌تر برای قطعه‌بندی سایر تصاویر پزشکی به کار گرفته شده بود، استفاده شده و با روش اول (روش ساده‌تر)، مقدار میانگین 0.59 برای ضریب Dice به دست آمد. سپس در روش دوم، با تغییر اندازه‌ی پنجره و اعمال پس‌پردازش، مقدار میانگین 0.77 برای ضریب Dice به دست آورده شد. این پژوهش در مقایسه با روش‌های ارائه شده در [۳]، [۵] و [۶]، که به ترتیب به ضریب Dice 0.74 ، 0.70 و 0.73 دست یافته‌اند، به نتایج بهتری رسیده است. همچنین، تاثیر افزودن داده‌ها با روش‌های متفاوت، مورد آزمایش قرار گرفته و نتایج مربوط به آن گزارش شده است.



شکل (۵) - نمونه‌هایی از برش‌های قطعه‌بندی شده توسط روش دوم برای پنج حجم مختلف، (الف)-(ه) تصاویر خام، (و)-(ی) ناحیه‌ی مشخص شده توسط متخصص، (ک)-(س) چپ ناحیه‌ی مشخص شده توسط الگوریتم پیشنهادی



شکل (۶) - مقایسه‌ی DSC دو روش ارائه شده، به ازای دوره‌های مختلف

جدول (۱) - نتایج افزودن داده‌ها

وارونگی	مقیاس تغییر اندازه	زوایای چرخش	DSC
✓			0.524
✓	2.4, 0.25, 0.5		0.476
✓	2.0/5		0.586
✓	4.0/25		0.392
x			0.138
x	0.5		0.12
x	2		0.148
x	2.0/5		0.166
x	0.25		0.76
x	4		0.174
x	4.0/25		0.134
x	4.2, 0.25, 0.5		0.174
x	270, 180, 90		0.196

۴-۴ - مقایسه‌ی نتایج روش‌های اول و دوم

در جدول (۲) نتایج به دست آمده از دو روش پیشنهادی ارائه شده است. در سطر اول، میانگین و انحراف معیار Dice برای روش اول، با اندازه‌ی پنجره‌ی ثابت و بدون پس‌پردازش، نشان داده شده است. چهار سطر بعدی جدول، این نتایج را برای روش دوم، با اندازه‌ی پنجره‌ها و مقیاس‌های کوچک‌نمایی شده‌ی متفاوت، نشان می‌دهند.

جدول (۲) - نتایج به دست آمده از روش‌های پیشنهادی

روش	اندازه‌ی پنجره	مقیاس‌های کوچک‌نمایی در پس‌پردازش	میانگین DSC	انحراف معیار
روش اول	$80 \times 80 \times 32$	بدون پس‌پردازش	0.59	0.27
روش دوم	$40 \times 40 \times 16$	0.5, 0.6, 0.7, 0.8, 0.9	0.77	0.16
		0.5, 0.75	0.74	0.18
		0.5, 0.6, 0.7, 0.8, 0.9	0.77	0.14
		0.4, 0.6, 0.8	0.75	0.17

در شکل (۵)، پنج نمونه از اسلایس‌های میانی حجم‌ها، که با روش پیشنهادی دوم قطعه‌بندی شده‌اند، ارائه شده است. همچنین، برای مقایسه‌ی بهتر روند پیش‌رفت دقت حین آموزش شبکه، نمودار ضریب Dice به ازای هر فولد و ایپاک در روش دوم، در شکل (۶) ارائه شده است. در این روش، مقدار 0.77 برای ضریب Dice به دست آمده است که بهبود چشم‌گیری را نسبت به روش قبل نشان می‌دهد.

۵ - نتیجه‌گیری

در این پژوهش، سعی بر آن شد تا قطعه‌بندی تصاویر سه‌بعدی ABUS، با کمک یادگیری ژرف، انجام گردد. قطعه‌بندی در این تصاویر، به معنای تشخیص نواحی توده از بافت طبیعی پستان است. در این پژوهش، از ایده‌ی شبکه‌ی یادگیری ژرف

- [10] Ciccek, Abdulkadir, Ahmed and Lienkamp, Soeren S and Brox, Thomas and Ronneberger, Olaf, "3D U-Net: learning dense volumetric segmentation from sparse annotation," International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 424-432, 2016.
- [11] Kingma, Diederik P and Ba, Jimmy, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

استفاده از معماری‌های جدید شبکه‌های عصبی ژرف، برای افزایش دقت قطعه‌بندی و نیز استفاده از این شبکه‌ها برای شناسایی و تشخیص نوع توده‌ها، به عنوان ادامه‌ی کار، مد نظر قرار داده شده است.

۶- مراجع

- [1] Jemal, Ahmedin and Bray, Freddie and Center, Melissa M and Ferlay, Jacques and Ward, Elizabeth and Forman, David "Global cancer statistics," CA: a cancer journal for clinicians, vol. 61, no.2, pp.69-90, 2011.
- [2] Cheng, Heng-Da, Juan Shan, Wen Ju, Yanhui Guo, and Ling Zhang. "Automated breast cancer detection and classification using ultrasound images: A survey." Pattern recognition 43, no. 1 (2010): 299-317.
- [3] [Kuo, H and Giger, Maryellen L and Reiser, Ingrid and Drukker, Karen and Edwards, Alexandra and Sennett, Charlene A, "Automatic 3D lesion segmentation on breast ultrasound images," Proc SPIE, vol. 8670, pp. 867025, 2013.
- [۴] ا. کوزه‌گر، م. سریانی، ح. بهنام، م. سلامتی، ت. تن، «تعیین مرز توده در تصاویر سه بعدی اولتراسوند خودکار با استفاده از یک مدل شکل‌پذیر» فصلنامه علمی - پژوهشی بیماری‌های پستان ایران، جلد ۱۰، شماره ۲، صفحه ۱۶-۲۶، ۱۳۹۶.
- [5] E. Kozegar and M. Soryani and H. Behnam and M. Salamati and T. Tan, "Mass Segmentation in Automated 3-D Breast Ultrasound Using Adaptive Region Growing and Supervised Edge-Based Deformable Model," IEEE Transactions on Medical Imaging, vol. PP, no. 99, pp. 1-1, 2017.
- [6] Tan, Tao and Gubern-Merida, Albert and Borelli, Cristina and Manniesing, Rashindra and van Zelst, Jan and Wang, Lei and Zhang, Wei and Platel, Bram and Mann, Ritse M and Karssemeijer, Nico, "Segmentation of malignant lesions in 3D breast ultrasound using a depth-dependent model," Medical physics, vol. 43, no. 7, pp. 4074-4084, 2016.
- [7] Wang, Jiahui and Engelmann, Roger and Li, Qiang, "Segmentation of pulmonary nodules in three-dimensional CT images by use of a spiral-scanning technique," Medical physics, vol. 34, no. 12, pp. 4678-4689, 2007
- [8] Ciresan, Dan and Giusti, Alessandro and Gambardella, Luca M and Schmidhuber, Jurgen, "Deep neural networks segment neuronal membranes in electron microscopy images," Advances in neural information processing systems, pp. 2843-2851, 2012.
- [9] Ronneberger, Olaf and Fischer, Philipp and Brox, Thomas, "U-net: Convolutional networks for biomedical image segmentation," International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234-241, 2015.