



دانشگاه صنعتی امیرکبیر

(پلی تکنیک تهران)

دانشکده مهندسی کامپیوتر

پروژه تحقیقاتی درس شبکه‌های عصبی

شبکه‌ی یو و بهبودهای آن

نگارش

عطیه غفارلوی مقدم

استاد درس

دکتر رضا صفابخش

تیر ۱۴۰۲

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

## چکیده

با توجه به رشد بی سابقه‌ی به کارگیری شبکه‌های عمیق در بینایی ماشین در دهه‌های گذشته، بسیاری از این شبکه‌ها برای تحلیل تصاویر پزشکی به کار گرفته شده‌اند. مساله‌ی قطعه‌بندی معنایی به معنی تقسیم تصویر به قطعات معنادار یک مساله‌ی مهم و پرچالش در حوزه‌ی تحلیل تصاویر پزشکی می‌باشد. شبکه‌ی یو یک رویکرد مبتنی بر یادگیری عمیق برای حل مساله‌ی قطعه‌بندی معنایی تصاویر پزشکی است که با توجه به عملکرد بسیار مناسب آن به خصوص با مجموعه داده‌ی کوچک به سرعت مورد توجه محققین قرار گرفت. اما علی رغم عملکرد بسیار خوب این شبکه نسبت به نسخه‌های پیش از خود، همچنان مشکلاتی در این شبکه وجود داشت که رفع آن‌ها باعث افزایش هر چه بیشتر قابلیت‌های این شبکه می‌شود. در این گزارش ابتدا شبکه‌ی یو ی اولیه را معرفی می‌کنیم و سپس با اشاره به هر یک از مشکلات این شبکه یک نسخه‌ی بهبود یافته‌ی شبکه که در صدد رفع آن مشکل برآمده است را معرفی و تفاوت‌ها و نتایج ارزیابی هر یک را بررسی می‌کنیم.

## واژه‌های کلیدی:

قطعه‌بندی معنایی، شبکه‌ی یو، یو-نت، ساختار کدگذار کدگشا

# فهرست مطالب

صفحه

عنوان

۱	.....	۱	مقدمه
۵	.....	۲	شبکه‌ی یو ..
۶	.....	۱-۲	معماری شبکه
۷	.....	۲-۲	آموزش شبکه
۸	.....	۳-۲	نتایج
۱۰	.....	۳	شبکه‌ی یو++ ..
۱۱	.....	۱-۳	معماری شبکه
۱۲	.....	۲-۳	آموزش شبکه
۱۳	.....	۳-۳	نتایج
۱۵	.....	۴	شبکه‌ی یو ۳+ ..
۱۶	.....	۱-۴	معماری شبکه
۱۸	.....	۲-۴	آموزش شبکه
۱۸	.....	۳-۴	نتایج
۲۰	.....	۵	شبکه‌ی یوی سبک و کارا ..
۲۱	.....	۱-۵	معماری شبکه
۲۲	.....	۲-۵	آموزش شبکه
۲۳	.....	۳-۵	نتایج
۲۴	.....	۶	شبکه‌ی یوی چند باقی‌مانده‌ای ..
۲۶	.....	۱-۶	معماری شبکه
۲۷	.....	۲-۶	آموزش شبکه
۲۸	.....	۳-۶	نتایج
۲۹	.....		کتاب‌نامه

شکل	فهرست تصاویر	صفحه
۱-۱	تنوع مقیاس در جراحی‌های پوستی در تصاویر درموسکوپی [۶]	۳
۱-۲	معماری شبکه‌ی یو [۹]	۷
۱-۳	معماری شبکه‌ی یو++ [۱۱]	۱۲
۱-۴	مقایسه‌ی اتصالات در سه شبکه‌ی یو، یو++ و یو+ [۵]	۱۶
۲-۴	چگونگی ساخت یک مجموعه ویژگی تمام مقیاس در شبکه‌ی یو+ [۵]	۱۷
۱-۵	مقایسه‌ی اتصالات در چهار شبکه‌ی یو، یو++، یو+ و یو سوک [۳]	۲۲
۲-۵	چگونگی ساخت یک مجموعه ویژگی تجمیع شده برای لایه‌ی دوم کدگشا [۳]	۲۳
۱-۶	ساخت مرحله به مرحله‌ی بلوک چند باقی‌مانده‌ای [۶]	۲۵
۲-۶	مسیر باقی‌مانده‌ای در شبکه‌ی یو چند باقی‌مانده‌ای [۶]	۲۶
۳-۶	معماری شبکه‌ی یو چند باقی‌مانده‌ای [۶]	۲۷

صفحه	فهرست جداول	جدول
۸	رتبه‌بندی چالش قطعه‌بندی $EM$ [۹]	۱-۲
۹	نتایج قطعه‌بندی چالش $ISBI$ [۹]	۲-۲
۱۴	نتایج قطعه‌بندی برای شبکه‌ی یو با و بدون نظارت عمیق [۱۱]	۱-۳
۱۹	نتایج ارزیابی شبکه‌ی یو+۳ با $Vgg - 16$ بر روی دو مجموعه داده‌ی طحال و کبد [۵]	۱-۴
۲۳	نتایج به دست آمده از شبکه‌ی یوی سوک در مقایسه با سایر شبکه‌ها بر روی مجموعه داده‌ی $ISBI$ [۳]	۱-۵
۲۸	مقایسه‌ی تعداد پارامترهای شبکه‌ی یو و شبکه‌ی یوی چند باقی‌مانده‌ای [۶]	۱-۶
۲۸	جدول نتایج ارزیابی شبکه‌ی یو چند باقی‌مانده‌ای روی پنج مجموعه داده‌ی تصاویر پزشکی [۶]	۲-۶

## فهرست اختصارات

عنوان اختصاری    عنوان کامل

شبکه‌ی یوی سوک    شبکه‌ی یوی سبک و کارا

# فصل اول

## مقدمه



از زمان پیدایش ابزارهای تصویربرداری پزشکی دیجیتال توجه زیادی روی پیاده‌سازی و به کارگیری تکنیک‌های پردازش تصویر در تحلیل این مجموعه تصاویر متمرکز شده است. مساله‌ی تشخیص پزشکی خودکار نیز خود متشکل از وظایف مختلف تحلیل تصویر از جمله قطعه‌بندی<sup>۱</sup> و تشخیص موجودیت هدف می‌باشد. در مساله‌ی قطعه‌بندی معنایی هدف تقسیم تصویر به تعدادی ناحیه‌ی ناهمپوشان است که پیکسل‌های متعلق به یک موجودیت را در یک خوشه قرار می‌دهد. در بحث قطعه‌بندی تصاویر پزشکی هدف اصلی پیدا کردن نواحی مورد نظر از جمله تومور، ارگان‌های بدن، آسیب دیدگی و غیره در انواع مختلفی از تصاویر از جمله ام آر آی و سی تی اسکن می‌باشد.

روش‌های اولیه‌ی ارائه شده برای حل مساله‌ی قطعه‌بندی در تصاویر پزشکی، رویکردهای ساده‌ی مبتنی بر قانون<sup>۲</sup> بودند که نسخه‌های پیشرفته‌تر آنها از الگوریتم‌های فازی<sup>۳</sup> و محاسبات نرم بهره می‌بردند. مشکل عمده‌ی این روش‌ها دخیل بودن نظر متخصص به تبع آن بایاس انسانی و عدم توانایی تطبیق با تنوع داده‌ی موجود در تصویربرداری پزشکی بود. در ادامه با گسترش شبکه‌های عمیق<sup>۴</sup> و شبکه‌های پیچشی<sup>۵</sup> رویکردهای حل مساله‌ی قطعه‌بندی به سمت روش‌های یادگیری عمیق متمایل شد. پیش از ارائه‌ی ساختارهای متناسب با مساله‌ی قطعه‌بندی معنایی، وظایف شبکه‌های پیچشی بیشتر در دسته‌بندی و برچسب‌گذاری کلیت یک تصویر خلاصه می‌شد؛ حال آنکه در قطعه‌بندی معنایی نیازمند برچسب‌دهی به هر پیکسل هستیم.

سیرسن و همکاران [۲] از یک رویکرد مبتنی بر پنجره‌ی لغزان<sup>۶</sup> برای به کارگیری شبکه‌های پیچشی در مساله‌ی قطعه‌بندی معنایی استفاده کردند. این رویکرد به واسطه‌ی افزونگی داده و سرعت آهسته‌ی آموزش شبکه‌چندان رویکرد کارایی محسوب نمی‌شد. به دنبال آن لانگ و همکاران [۸] یک شبکه‌ی تمام پیچشی<sup>۷</sup> ارائه کردند که اولین قطعه‌بندی تصویر به تصویر را انجام می‌داد. این شبکه عملکرد مناسبی داشت و پایه‌ی شبکه‌های بعدی خود از جمله شبکه‌ی سگ<sup>۸</sup> [۱] و شبکه‌ی یو [۹] گردید.

مشکل اصلی شبکه‌های تمام پیچشی نیازمندی آنها به داده‌ی آموزشی زیاد است و این مساله به طور خاص در مسائل پزشکی که جمع‌آوری داده و برچسب‌گذاری آن کار پرهزینه، زمانبر و نیازمند دقت بسیاری است؛ موجب عدم کارایی شبکه می‌گردد. شبکه‌ی یو در شکل اولیه‌ی خود برای حل این چالش در کنار افزایش دقت قطعه‌بندی ارائه شد. این شبکه دارای یک ساختار کدگذار-کدگشا<sup>۹</sup> می‌باشد

<sup>1</sup>segmentation

<sup>2</sup>rule-based

<sup>3</sup>fuzzy algorithms

<sup>4</sup>deep networks

<sup>5</sup>convolutional networks

<sup>6</sup>sliding window

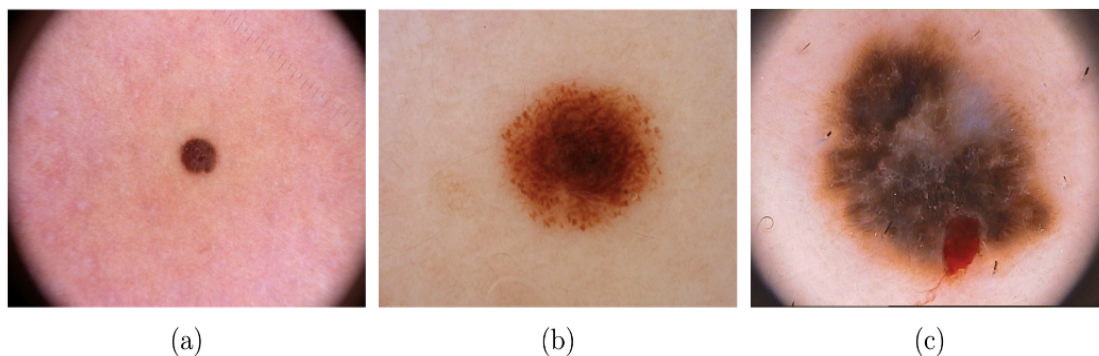
<sup>7</sup>fully convolutional network

<sup>8</sup>segNet

<sup>9</sup>encoder-decoder

که در آن لایه‌های متناظر در کدگذار و کدگشا با اتصالات پرشی<sup>۱۰</sup> به یکدیگر متصل هستند. شبکه‌ی یو انقلابی در زمینه‌ی قطعه بندی معنایی تصاویر پزشکی به وجود آورد و بعدها در حوزه‌های مختلف و برای تصاویر و ساختارهای هدف دیگری چون ساختارهای عصبی، کلیه، بافت های پوستی و غیره مورد استفاده قرار گرفت.

با وجود عملکرد اولیه‌ی مناسب شبکه‌ی یو دو مشکل عمده در این شبکه وجود داشت که بهبودهای بعدی در تلاش برای حل این دو مشکل ارائه شدند. اولین مشکل یک شبکه‌ی یو عدم مقاومت آن در برابر تغییر مقیاس می‌باشد. در مسائل پزشکی ما با یک موجودیت با اندازه و مقیاس ثابت و مرز های مشخص طرف نیستیم و مرز قطعات و اندازه ی آن‌ها می‌تواند بسیار متنوع و نامشخص باشد. شکل ۱-۱ تنوع این مقیاس ها در جراحات پوستی را نشان می‌دهد. مشکل دوم شبکه‌ی یو وجود فاصله ی معنایی<sup>۱۱</sup> میان اطلاعات استخراج شده در کدگذار و کدگشا می‌باشد. همانطور که پیش تر هم مطرح شد شبکه‌ی یو از اتصالات پرشی میان کدگذار و کدگشا استفاده می‌کند و با توجه به اینکه کدگذار اطلاعات سطح پایین و جزئیات را دارد و کدگشا اطلاعات سطح بالا و معنایی را، بین اطلاعات ترکیب شده یک فاصله‌ی معنایی وجود دارد که می‌تواند کارایی شبکه را کاهش دهد.



شکل ۱-۱: تنوع مقیاس در جراحات های پوستی در تصاویر درموسکوپي [۶]

در این گزارش هدف بررسی معماری اولیه‌ی شبکه‌ی یو و بهبودهای اخیر آن از جمله شبکه‌ی یو++ [۱۱]، شبکه‌ی یو ۳+ [۵]، شبکه‌ی یوی سبک و کارا<sup>۱۲</sup> [۳]، شبکه‌ی یوی چندباقیمانده‌ای<sup>۱۳</sup> [۶] می‌باشد. هر کدام از این بهبودها ناظر به برطرف کردن یک یا هر دو مشکل مطرح شده در بند قبل هستند و این بهبود از طریق تغییر در معماری و تابع خطا ایجاد شده است. در ادامه‌ی گزارش برای هر شبکه یک فصل در نظر گرفته شده است که شامل توضیحات کلی شبکه، ساختار و معماری شبکه و روش آموزش

<sup>10</sup>skip connections

<sup>11</sup>semantic gap

<sup>12</sup>ELU-net

<sup>13</sup>multiRes U-net

آن می‌باشد. با توجه به اینکه هدف اصلی در گزارش شناخت این بهبودها و مزیت هر کدام نسبت به سایر شبکه‌هاست؛ از پرداختن دقیق به نتایج متعدد به دست آمده روی مجموعه داده‌های مختلف پرهیز شده است و فقط به اندازه‌ای که نمایانگر مزیت شبکه‌ها نسبت به یکدیگر باشد به آن‌ها پرداخته شده است.

## فصل دوم

### شبکه‌ی یو

در این فصل قرار است شبکه‌ی یو را از سه جنبه‌ی معماری، آموزش و نتایج به دست آمده بررسی کنیم.

## ۱-۲ معماری شبکه

ساختار پایه‌ی یک شبکه‌ی یو از دو مسیر تشکیل شده است. مسیر اول مسیر انقباضی<sup>۱</sup> می‌باشد که به آن مسیر تحلیل<sup>۲</sup> یا کدگذار<sup>۳</sup> نیز گفته می‌شود. این مسیر شبیه به یک شبکه‌ی کانولوشنی معمولی مورد استفاده برای دسته بندی می‌باشد. مسیر دوم مسیر گسترش<sup>۴</sup>، کدگشا یا مسیر ساخت<sup>۵</sup> نیز نامیده می‌شود. در این مسیر از نمونه افزایشی<sup>۶</sup> و الحاق ویژگی‌هایی که به واسطه اتصالات پرشی به کدگشا رسیده‌اند مجموعه ویژگی‌های مورد نیاز ساخته می‌شوند. این گسترش به شبکه اجازه می‌دهد که اطلاعات محلی را در کنار اطلاعات سطح بالای استخراج شده برای قطعه‌بندی به کار گیرد و همچنین رزولوشن خروجی تولید شده را افزایش دهد و در نهایت یک تصویر تماماً قطعه‌بندی شده تولید کند. شبکه‌ی حاصل یک شبکه‌ی متقارن با ساختار یو شکل است که در شکل ۱-۲ نشان داده شده است.

همان طور که در شکل ۱-۲ نشان داده شده است هر بلوک کدگذار شامل دو کانولوشن  $3 \times 3$  متوالی و به دنبال آن یک تابع فعال‌سازی  $ReLU$  و سپس یک لایه‌ی ادغام حداکثر<sup>۷</sup> می‌باشد و این بلوک در چندین سطح تکرار می‌شود. در هر مرحله‌ی کدگذار تعداد صفحات ویژگی دو برابر می‌شود. ابداع اصلی شبکه‌ی یو در ساختار کدگشاست. در کدگشا هر بلوک شامل نمونه افزایشی با استفاده از کانولوشن‌های فزاینده<sup>۸</sup> با اندازه‌ی  $2 \times 2$  است و سپس مجموعه ویژگی لایه‌ی متناظر در کدگذار به مجموعه ویژگی به دست آمده الحاق<sup>۹</sup> می‌شود و سپس یک کانولوشن  $3 \times 3$  و یک تابع فعال‌سازی  $ReLU$  اعمال می‌گردد. در هر مرحله از کدگشا تعداد صفحات ویژگی نصف می‌شود و در نهایت در لایه‌ی آخر یک کانولوشن  $1 \times 1$  برای کاهش تعداد صفحات ویژگی اعمال می‌شود.

<sup>1</sup>contacting path

<sup>2</sup>analysis path

<sup>3</sup>encoder

<sup>4</sup>expansive path

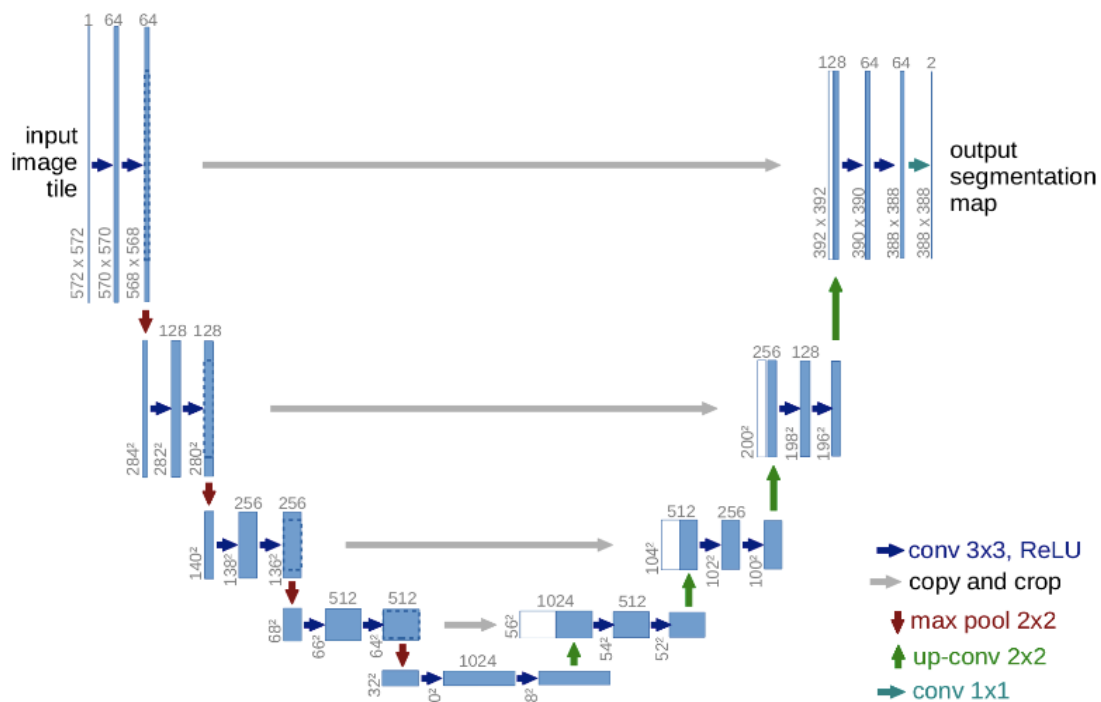
<sup>5</sup>synthesis path

<sup>6</sup>upsampling

<sup>7</sup>max pooling

<sup>8</sup>up-convolution

<sup>9</sup>concat



شکل ۲-۱: معماری شبکه‌ی یو [۹]

## ۲-۲ آموزش شبکه

آموزش شبکه با روش نزول در جهت گرادیان تصادفی<sup>۱۰</sup> صورت می‌گیرد. در فاز آموزش شبکه از داده افزایشی<sup>۱۱</sup> به صورت ایجاد تخریب و تغییر در تصاویر موجود استفاده شده است. این کار به شبکه اجازه می‌دهد که نسبت به اعوجاج و تغییرات موجود در تصاویر پزشکی مقاوم گردد. این موضوع در قطعه‌بندی معنایی تصاویر پزشکی بسیار حائز اهمیت است چرا که این گونه اعوجاج و تخریب و حتی کشیدگی تصویر در تصاویر پزشکی به کرات موجود است و شبکه باید در برابر آنها مقاوم باشد. تابع هدف این شبکه به صورت زیر می‌باشد:

$$E = \sum_{x \in \Omega} w(x) \log(p_{l(x)}(X)) \quad (۱-۲)$$

این تابع از اعمال یک لایه ی سافت‌مکس بر روی مجموعه ویژگی و ترکیب آن با خطای آنتروپی

<sup>10</sup>stochastic gradient descent

<sup>11</sup>data augmentation

<i>rank</i>	<i>group name</i>	<i>warping error</i>	<i>rand error</i>	<i>pixel error</i>
	<i>** human values **</i>	0.000005	0.0021	0.0010
1.	<i>u - net</i>	0.000353	0.0382	0.0611
2.	<i>DIVE - SCI</i>	0.000355	0.0305	0.0584
3.	<i>IDSIA</i>	0.000420	0.0504	0.0613
4.	<i>DIVE</i>	0.000430	0.0545	0.0582

جدول ۱-۲: رتبه‌بندی چالش قطعه‌بندی  $EM$  [۹]

متقابل<sup>۱۲</sup> به دست می‌آید و بنابراین  $p_k$  به صورت زیر قابل تعریف می‌باشد:

$$p_k(x) = \frac{\exp(a_k(x))}{\sum_{k'=1}^K \exp(a_{k'}(x))} \quad (۲-۲)$$

که  $a_k$  تابع فعال‌سازی کانال  $k$  ام است.

در شبکه‌های عمیق با تعداد زیادی لایه‌ی کانولوشنی و مسیرهای متعدد مقداردی‌اولیه به وزن‌ها امر بسیار مهمی محسوب می‌شود چرا که موجب فعال‌سازی بیش از حد بخش‌هایی از شبکه و مردن بخش‌های دیگر آن می‌گردد. در حالت ایده آل وزن دهی باید گونه ای باشد که هر صفحه‌ی ویژگی ساخته شده در شبکه تقریباً واریانسی نزدیک به یک داشته باشد. در شبکه‌ی یو برای محقق شدن این هدف وزن‌های اولیه از یک توزیع گاوسی با انحراف معیار  $\sqrt{\frac{2}{N}}$  آمده‌اند که  $N$  تعداد ورودی‌های یک نورون است.

## ۳-۲ نتایج

شبکه‌ی یو در دو وظیفه‌ی قطعه‌بندی تصاویر پزشکی مورد ارزیابی قرار گرفته است. اولین مجموعه داده مربوط به داده‌های چالش  $EM$  است که شامل ۳۰ تصویر ساختارهای عصبی به همراه ماسک‌های قطعه‌بندی واقعی می‌باشد. جدول ۱-۲ عملکرد شبکه‌ی یو در این چالش را نشان می‌دهد. و مجموعه داده‌ی دوم متعلق به چالش قطعه‌بندی سلول‌ها در تصاویر میکروسکوپ نوری است که به چالش  $ISBI$  معروف است. جدول ۲-۲ نتایج شبکه‌ی یو در این چالش را نشان می‌دهد. همان طور که در جداول نیز قابل مشاهده است؛ شبکه‌ی یو در هر دو چالش موفق به کسب بهترین عملکرد شده است.

<sup>12</sup>cross entropy loss function

<i>name</i>	<i>PhC – U373</i>	<i>DIC – HeLa</i>
<i>IMCB – SG</i> (2014)	0.2669	0.2935
<i>KTH – SE</i> (2014)	0.7953	0.4607
<i>HOUS – US</i> (2014)	0.5323	–
<i>second – best</i> (2015)	0.83	0.46
<i>u – net</i> (2015)	0.9203	0.7756

جدول ۲-۲: نتایج قطعه‌بندی چالش *ISBI* [۹]



## فصل سوم

### شبکه‌ی یو++

در بحث قطعه‌بندی معنایی در حوزه‌ی تصویر، تشخیص اشتباه مرزها می‌تواند منجر به تشخیص اشتباه به طور خاص در مورد خوش‌خیم یا بدخیم بودن یک بیماری شود؛ از این رو دقت این شبکه‌های قطعه‌بندی یک مسأله‌ی غیرقابل مصالحه است. با وجود عملکرد مناسبی که شبکه‌ی یو نسبت به رقبا خود در چالش‌های پزشکی داشت، همچنان دقت این شبکه در تعیین برخی مرزها و مجموعه داده‌ها ضعیف بود. شبکه‌ی یو++ با این هدف توسعه داده شد که با برطرف کردن مشکلات یادگیری در شبکه‌ی یو دقت آن را افزایش دهد. ایده‌ی اصلی این شبکه این بود که اگر ویژگی‌های کدگذار که در فرآیند کدگذاری به کدگشا الحاق می‌شوند از نظر معنایی مشابه ویژگی‌های تولید شده در کدگشا باشند، در این صورت آموزش شبکه به مراتب آسانتر و دقت آن به مراتب بالاتر خواهد بود. در شبکه‌های پیش از شبکه‌ی یو++ مانند شبکه‌های تمام پیچشی و شبکه‌ی یو ویژگی‌های کدگذار بدون اعمال هیچ تغییری به ویژگی‌های کدگشا الحاق می‌شدند اما در شبکه‌ی یو++ این اتصالات دارای ساختارهای پیچیده تری هستند که فاصله‌ی معنایی میان ویژگی‌های دو زیرشبکه را کاهش می‌دهد.

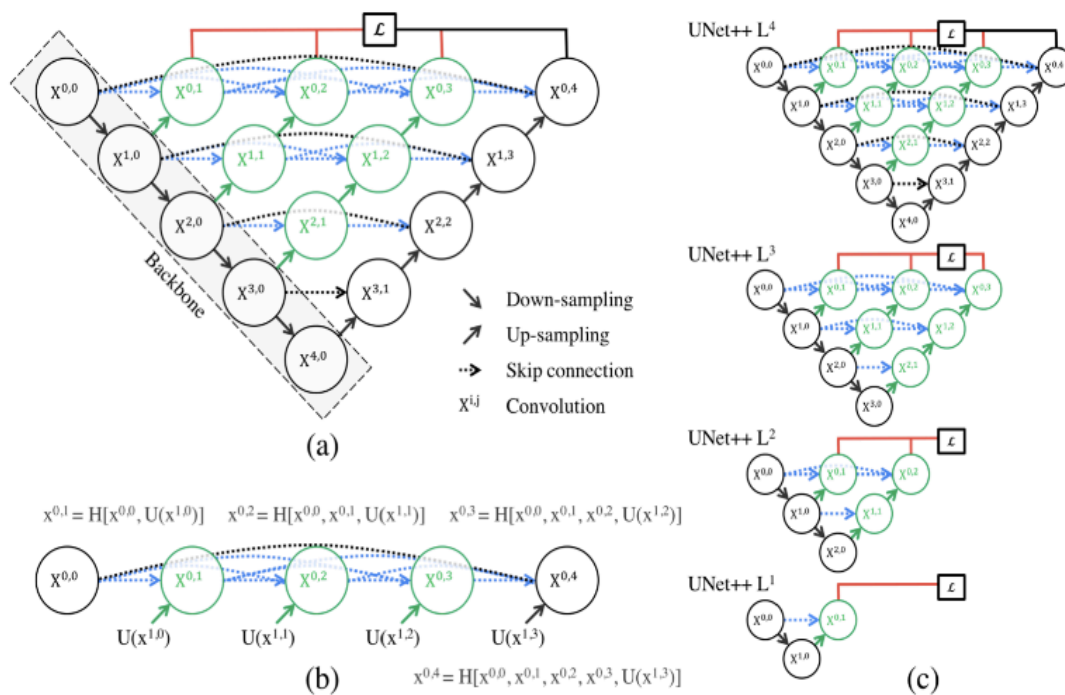
### ۱-۳ معماری شبکه

مشابه شبکه‌ی یو در این شبکه هم دو زیرشبکه‌ی کدگذار و کدگشا داریم. تفاوت اصلی در اتصالات پرشی میان ویژگی‌های کدگذار و کدگشا است. در شبکه‌ی یو++ صفحات ویژگی کدگذار از یک بلوک کانولوشنی فشرده<sup>۱</sup> عبور می‌کنند که تعداد لایه‌های آن بسته به اینکه اتصال در کدام سطح از شبکه قرار دارد متفاوت است. کار اصلی این بلوک کانولوشنی فشرده هم‌سطح کردن اطلاعات معنایی تولید شده توسط کدگذار و اطلاعات خروجی کدگشا می‌باشد. در این شبکه می‌توان نشان داد که با مشابه کردن این اطلاعات، بهینه‌سازی شبکه بسیار ساده‌تر انجام خواهد شد. شکل ۱-۳ قسمت a ساختار شبکه و مسیرهای پرشی اضافه شده به آن را نشان می‌دهد. همان طور که در این شکل هم مشاهده می‌شود تعداد لایه‌های بلوک کانولوشنی فشرده در هر اتصال به سطح آن اتصال در هرم شبکه‌ی یو++ وابسته است. به طور کلی می‌توان مسیرهای پرشی در یک شبکه را به این صورت تعریف کرد که فرض کنید  $x_{i,j}$  خروجی گره ی  $X_{i,j}$  است که  $i$  درایه‌ی لایه‌ی کدگذار و  $j$  درایه‌ی لایه‌ی کانولوشنی بلوک فشرده در مسیر اتصال پرشی باشد؛ در این صورت مجموعه صفحات ویژگی به دست آمده برای  $x_{i,j}$  به صورت زیر می‌باشد:

$$x_{i,j} = \begin{cases} \mathcal{H}(x_{i-1,j}) & j = 0 \\ \mathcal{H}([x_{i,k}]_{k=0}^{j-1}, \mathcal{U}(x_{i+1,j-1})) & j > 0 \end{cases} \quad (1-3)$$

که  $\mathcal{H}(\cdot)$  عملگر کانولوشن به همراه تابع فعال‌سازی است و  $\mathcal{U}(\cdot)$  یک لایه‌ی نمونه‌افزایی در کدگشا و  $[]$  نیز نماد الحاق می‌باشد.

<sup>1</sup>dense convolution block



شکل ۳-۱: معماری شبکه‌ی یو++ [۱۱]

## ۲-۳ آموزش شبکه

برای آموزش شبکه‌ی یو++ از رویکرد نظارت عمیق<sup>۲</sup> استفاده می‌شود [۷]. نظارت عمیق یک رویکرد آموزش در شبکه‌های عمیق است که لایه‌های میانی هم علاوه بر لایه‌ی نهایی ارزیابی می‌شوند و توابع زیان مختص به خود را دارند. این توابع اضافی تعریف شده بازخورد بیشتری برای آموزش شبکه تولید می‌کنند و می‌توانند باعث همگرایی سریعتر آن شوند. نظارت عمیق در شبکه‌ی یو++ به دو صورت اعمال می‌گردد:

- حالت صحیح<sup>۳</sup>: در این حالت نقشه‌ی قطعه‌بندی نهایی از میانگین همه‌ی خروجی‌های قطعه‌بندی تولید می‌شود.
- حالت سریع<sup>۴</sup>: در این حالت نقشه‌ی قطعه‌بندی نهایی فقط از یکی از خروجی‌های شاخه‌های قطعه‌بندی تولید می‌شود که بسته به این که کدام شاخه انتخاب شود می‌تواند موجب هرس شدن درخت و افزایش سرعت آموزش گردد. شکل ۳-۱ قسمت c این موضوع را به خوبی نشان می‌دهد که بسته به نوع انتخابی که برای نظارت عمیق در حالت سریع داریم شبکه‌ی حاصل هرس شده

<sup>2</sup>deep supervision

<sup>3</sup>accurate

<sup>4</sup>fast mode

و سرعت یادگیری افزایش پیدا می‌کند.

با توجه به وجود مسیرهای پرشی در شبکه‌ی یو++ این شبکه قابلیت تولید صفحات ویژگی با رزولوشن کامل را در سطوح مختلف معنایی دارد و هر کدام از این سطوح و یا ترکیب آنها می‌تواند برای نظارت عمیق به کار گرفته شود.

تابع خطا در این شبکه ترکیبی از آنتروپی متقابل دودویی<sup>۵</sup> و ضریب تشابه<sup>۶</sup> برای هر یک از چهار سطح معنایی می‌باشد که به صورت زیر قابل تعریف است:

$$\mathcal{L}(Y, \hat{Y}) = -\frac{1}{N} \sum_{b=1}^N \left( \frac{1}{2} \cdot Y_b \cdot \log \hat{Y}_b + \frac{2 \cdot Y_b \cdot \hat{Y}_b}{Y_b + \hat{Y}_b} \right) \quad (2-3)$$

که  $Y_b$  و  $\hat{Y}_b$  به ترتیب احتمالات پیش بینی شده و مقادیر واقعی تصویر  $b$  ام هستند و  $N$  هم اندازه ی دسته‌ی<sup>۷</sup> داده هاست.

### ۳-۳ نتایج

آزمایش شبکه‌ی یو++ روی چهار مجموعه داده‌ی تصاویر پزشکی انجام شده است و مدل‌های پایه‌ی آزمایش، شبکه‌ی یو و شبکه‌ی یوی گسترده<sup>۸</sup> هستند. شبکه‌ی یوی گسترده در واقع شکلی از شبکه‌ی یوی اولیه است که تعداد پارامترهای آن با پارامترهای شبکه‌ی یو++ مساوی است. این مقایسه از این رو صورت گرفته که نشان داده شود که برتری شبکه‌ی یو++ به دلیل افزایش چشمگیر تعداد پارامترهای آن نیست. نتایج پیاده سازی در جدول ۱-۳ آورده شده است که تعداد پارامترهای هر کدام و میزان عملکرد آنها را در چهار مجموعه داده نشان می‌دهد. همانطور که مشاهده می‌کنید شبکه‌ی یو در هر دو نسخه ی بدون نظارت عمیق و با نظارت عمیق عملکرد بهتری نسبت به سایر شبکه‌ها دارد.

<sup>5</sup>binary cross entropy

<sup>6</sup>Dice coefficient

<sup>7</sup>batch

<sup>8</sup>wide u-net

<i>Architecture</i>	<i>params</i>	<i>cell nuclei</i>	<i>colon polyp</i>	<i>liver</i>	<i>lung nodule</i>
<i>U – net</i>	7.76M	90.77	30.08	76.62	71.47
<i>Wide U – net</i>	9.13M	90.92	30.14	76.58	73.38
<i>U – net ++ w/o DS</i>	9.04M	92.63	33.45	79.70	76.44
<i>U – net ++ w DS</i>	9.04M	92.52	32.12	82.90	77.21

جدول ۳-۱: نتایج قطعه‌بندی برای شبکه‌ی یو با و بدون نظارت عمیق [۱۱]

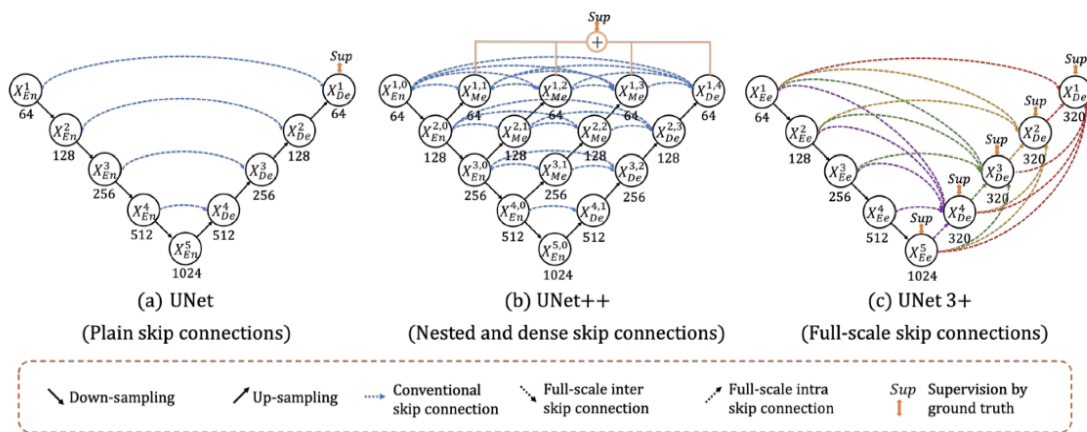
## فصل چهارم

### شبکه‌ی یو ۳+

در فصل قبلی گفتیم که مساله‌ی دقت در بحث قطعه‌بندی تصاویر پزشکی مساله‌ای بسیار مهم است. در شبکه‌ی یو++ با وجود اینکه قدرت قطعه‌بندی شبکه نسبت به شبکه‌ی یو افزایش پیدا کرده است اما این قدرت با هزینه‌ی صرف پارامترهای زیاد به دست آمده است. در شبکه‌ی یو ۳+ هدف افزایش قدرت در کنار کاهش تعداد پارامترهاست. مشکل دیگری که در شبکه‌ی یو++ وجود دارد این است که این شبکه قدرت استخراج اطلاعات کافی از تمام مقیاس‌ها را ندارد. هر صفحه‌ی ویژگی در یک مقیاس مشخص اطلاعات تمایزگر را استخراج می‌کند مثلاً صفحات ویژگی سطح پایین جزئیات بیشتر و ویژگی‌های مکانی را به خوبی استخراج می‌کنند که برای تعیین مرزهای موجودیت‌های تصویر مناسب است. در مقابل صفحات ویژگی سطح بالا اطلاعات معنایی در کلیت تصویر را به خوبی استخراج می‌کنند. ترکیب اطلاعات سطوح مختلف به ما یک مجموعه اطلاعات تمام مقیاس<sup>۱</sup> می‌دهد و هدف شبکه‌ی یو ۳+ هم استخراج این اطلاعات تمام مقیاس است.

## ۱-۴ معماری شبکه

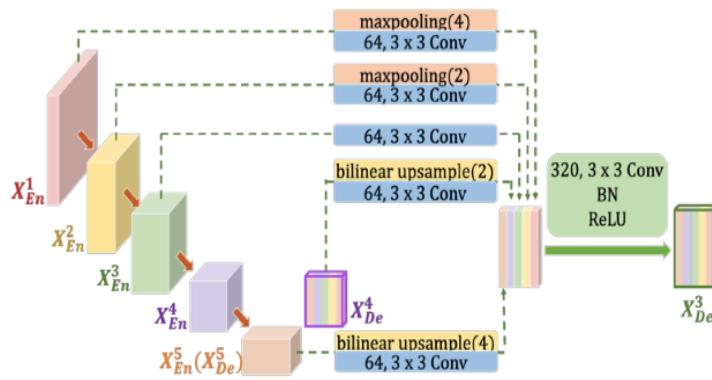
در این شبکه مانند شبکه‌ی یو اتصالات بین کدگذار و کدگشا در سطوح متناظر وجود دارد ولی برای این که یک استخراج تمام مقیاس صورت گیرد اتصالات پرشی بیشتری نیز از لایه‌های قبلی کدگذار و کدگشا به لایه‌ی مد نظر وجود دارد. در این شبکه، هر لایه‌ی کدگشا صفحات ویژگی با مقیاس کوچک‌تر و یا مقیاس یکسان را از کدگذار و همچنین صفحات ویژگی با مقیاس بزرگ‌تر را از کدگشا دریافت می‌کند. این امر باعث می‌شود که جزئیات و ویژگی‌های معنایی را در تمام مقیاس‌ها داشته باشد. شکل ۱-۴ تفاوت اصلی اتصالات شبکه‌ی یو ۳+ را با شبکه‌ی یو و یو++ نشان می‌دهد.



شکل ۱-۴: مقایسه‌ی اتصالات در سه شبکه‌ی یو، یو++ و یو ۳+ [۵]

<sup>۱</sup>full scale

شکل ۲-۴ چگونگی ساخت یک مجموعه ویژگی تمام مقیاس را برای لایه‌ی سوم کدگشا نشان می‌دهد. ابتدا شبیه به شبکه‌ی یو صفحات و ویژگی متناظر در کدگذار لایه‌ی سوم  $X_{En}^3$  با عبور از یک کانولوشن  $3 \times 3$  دریافت می‌شوند. همچنین تعدادی اتصال پرشی از لایه‌های کدگذار با سطح کوچکتر نیز دریافت می‌شوند، در این مثال این اتصالات از سطوح  $X_{En}^1$  و  $X_{En}^2$  متصل می‌شوند. این ویژگی‌ها پس از عبور از یک کانولوشن  $3 \times 3$  و یک لایه‌ی ادغام حداکثر در کدگشا دریافت می‌شوند. این مجموعه ویژگی شامل اطلاعات جزئی و سطح پایین است و وظیفه‌ی تامین ویژگی‌های با مقیاس کوچکتر را بر عهده دارد. و در نهایت یک مجموعه اتصالات داخل کدگشا و از لایه‌های با سطح بزرگتر دریافت می‌شوند. در این مثال این اطلاعات از سطوح  $X_{De}^4$  و  $X_{De}^5$  و پس از عبور از یک کانولوشن  $3 \times 3$  و یک نمونه‌افزایی دوطرفی<sup>۲</sup> در لایه‌ی سوم دریافت می‌شوند. وظیفه‌ی این ویژگی‌ها دخیل کردن اطلاعات سطح بالای معنایی و ویژگی‌های با مقیاس بزرگتر می‌باشد. از الحاق این ۵ مجموعه ویژگی و حذف اطلاعات زائد آن‌ها با عبور از یک کانولوشن  $3 \times 3$  و نرمال‌سازی دسته‌ای<sup>۳</sup> و گذر از تابع فعال‌سازی  $ReLU$  یک مجموعه ویژگی تمام مقیاس برای لایه‌ی سوم کدگشا مهیا می‌گردد.



شکل ۲-۴: چگونگی ساخت یک مجموعه ویژگی تمام مقیاس در شبکه‌ی یو ۳+ [۵]

اتصالات پرشی را میتوان با رابطه‌ی ۱-۴ نمایش داد. فرض کنید  $i$  درایه‌ی لایه‌ی کدگذار باشد و  $N$  تعداد کل لایه‌های کدگذار باشد. در این صورت مجموعه ویژگی تولید شده برای  $X_{De}^i$  به صورت زیر است:

$$X_{De}^i = \begin{cases} X_{En}^i & i = N \\ \mathcal{H}([\mathcal{C}(\mathcal{D}(X_{En}^k))_{k=1}^{i-1}, \mathcal{C}(\mathcal{U}(X_{De}^k))_{k=i+1}^N]) & i = 1, \dots, N-1 \end{cases} \quad (1-4)$$

<sup>۲</sup>bilinear upsampling

<sup>۳</sup>batch normalization



که تابع  $\mathcal{C}(\cdot)$  یک عملگر کانولوشنی است و  $\mathcal{H}(\cdot)$  شامل یک عملگر کانولوشن، یک نرمال‌سازی دسته‌ای و یک تابع فعال‌سازی  $ReLU$  است که در واقع همان رویکرد ترکیبی پس از الحاق ویژگی‌هاست و  $\mathcal{D}(\cdot)$  عملگر کاهش نمونه<sup>۴</sup> و  $\mathcal{U}(\cdot)$  عملگر نمونه‌افزایی هستند. شبکه‌ی ۳+ از تعداد پارامترهای کمتری نسبت به دو شبکه‌ی یو و یو++ برخوردار است و همین امر در کنار دقت مناسب آن این شبکه را به شبکه‌ی مناسبی برای قطعه بندی تبدیل می‌کند.

## ۲-۴ آموزش شبکه

در آموزش شبکه‌ی یو ۳+ نیز مانند شبکه‌ی یو++ از نظارت عمیق کمک گرفته می‌شود. در شبکه‌ی یو++ نظارت عمیق فقط در سطح رزولوشن کامل ( $X^{0,j}$ ) انجام می‌شد ولی در شبکه‌ی یو ۳+ خروجی هر سطح از کدگشا بر اساس حقیقت مبنا<sup>۵</sup> مورد ارزیابی قرار می‌گیرد. برای انجام نظارت عمیق در هر لایه ی کدگشا، ابتدا خروجی آخرین لایه از یک کانولوشن  $3 \times 3$  عبور می‌کند و سپس یک نمونه افزایشی دوخطی و یک تابع سیگموئید بر روی آن اعمال می‌گردد. در این شبکه از یک تابع خطای چند بخشی استفاده می‌شود که هر بخش وظیفه‌ی اعمال خطای قطعه‌بندی در یک سطح از سلسله مراتب پیکسل، قطعه<sup>۶</sup> و نقشه را انجام می‌دهد. این تابع خطا شبکه را توانمند می‌سازد تا هم ویژگی‌های بزرگ مقیاس و هم مرزهای مشخص و واضح را استخراج کند.

## ۳-۴ نتایج

این شبکه بر روی دو مجموعه داده‌ی کبد و طحال مورد ارزیابی قرار گرفته است. مدل‌های پایه برای مقایسه، دو مدل یو و یو++ بوده‌اند و معیار ارزیابی مدل‌ها ضریب تشابه سرنسن-دایس<sup>۷</sup> است که به صورت زیر قابل تعریف می‌باشد:

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

که  $TP$  همان مثبت صادق،  $FN$  منفی کاذب و  $FP$  مثبت کاذب می‌باشد. جدول ۴-۱ نتایج به دست آمده را نشان می‌دهد. همانطور که در این جدول مشاهده می‌شود، عملکرد شبکه‌ی یو ۳+ در هر دو مجموعه داده و به صورت میانگین از دو شبکه‌ی یو و یو++ بهتر است و این در حالی‌ست که تعداد پارامترهای آن نیز کمتر است و همین امر کارایی این شبکه را نشان می‌دهد.

<sup>4</sup>downsampling

<sup>5</sup>ground truth

<sup>6</sup>patch

<sup>7</sup>Dice coefficient

<i>Architecture</i>	<i>params</i>	<i>Dice<sub>liver</sub></i>	<i>Dice<sub>spleen</sub></i>	<i>Dice<sub>average</sub></i>
<i>U – net</i>	39.39M	0.9206	0.9023	0.9237
<i>U – net + +</i>	47.18M	0.9278	0.9230	0.9352
<i>U – net3+ w/o DS</i>	26.97M	0.9489	0.9437	0.9511
<i>U – net3+</i>	26.97M	0.9550	0.9496	0.9552

جدول ۴-۱: نتایج ارزیابی شبکه‌ی یو ۳+ با  $Vgg - 16$  بر روی دو مجموعه داده‌ی طحال و کبد [۵]

## فصل پنجم

### شبکه‌ی یوی سبک و کارا

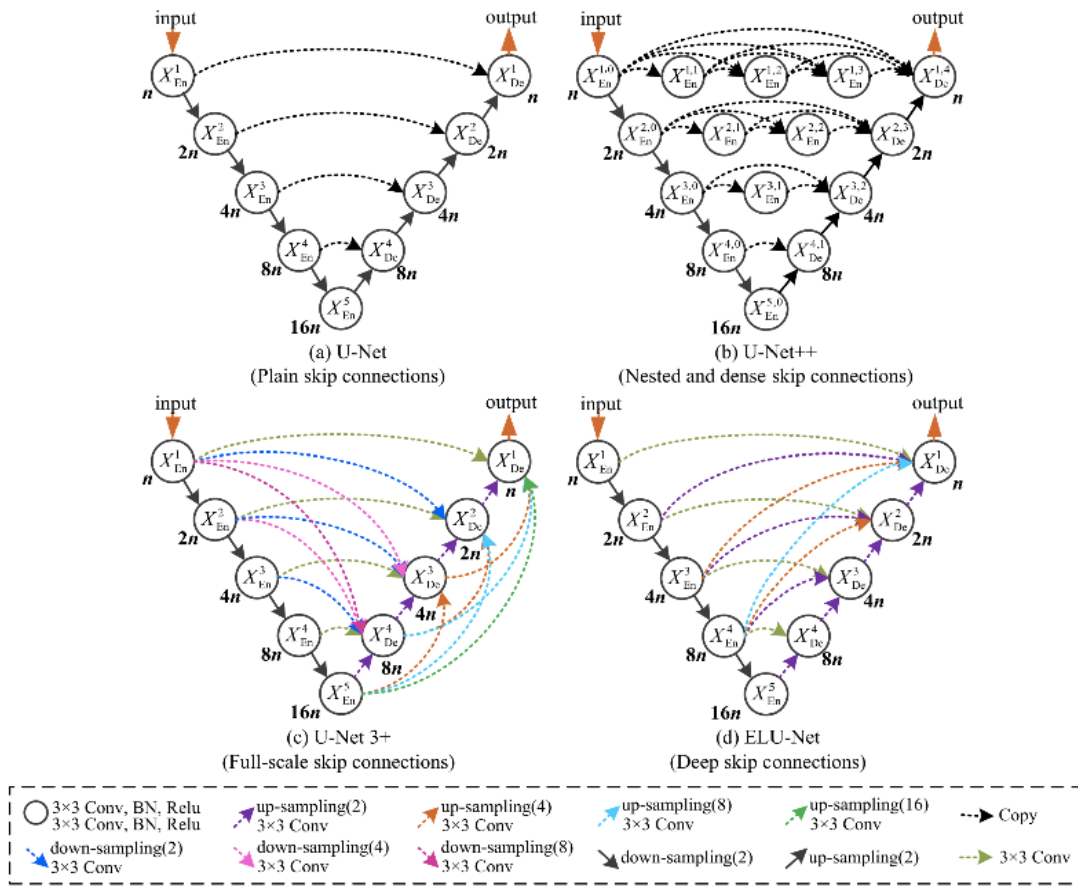
در شبکه‌ی یوی سبک و کارا، که از این پس به اختصار شبکه‌ی یوی سوک خوانده می‌شود، هدف مانند شبکه‌ی یو<sup>۳</sup>+ استخراج اطلاعات تمام مقیاس است. به طور خاص تمرکز شبکه‌ی یوی سوک روی استخراج اطلاعات تمام مقیاس از کدگذار می‌باشد. سازندگان این شبکه بر این باورند که با وجود عملکرد مناسب شبکه‌ی یو<sup>۳</sup>+ در کنار پارامترهای محدود آن این شبکه از تمام قابلیت‌هایی که یک شبکه‌ی یو مانند دارد بهره نمی‌گیرد [۳] و همچنان می‌توان در عین پایین نگه داشتن تعداد پارامترها مکانیزمی طراحی کرد که اطلاعات تمام مقیاس بیشتری استخراج کند. و بنابراین هدف از توسعه‌ی شبکه‌ی یوی سوک سه موضوع است:

- استفاده‌ی تمام و کمال از ویژگی‌های تمام مقیاس در کدگذار به وسیله‌ی تعریف اتصالات پرشی عمیق که از صفحات ویژگی با مقیاس بزرگ‌تر یا مساوی در کدگذار استفاده می‌کنند.
- تعریف بهینه‌ی تابع خطا به گونه‌ای که بیشترین یادگیری را برای شبکه ممکن سازد.
- به کارگیری کمترین تعداد پارامتر بدون افت عملکرد در چالش‌های حوزه‌ی پزشکی.

## ۵-۱ معماری شبکه

مشابه تمام ساختارهای شبکه‌ی یو در این شبکه نیز یک ساختار یو مانند حاوی دو زیر شبکه‌ی کدگذار و کدگشا وجود دارد. تفاوت اصلی آن مانند نمونه‌های قبلی در اتصالات پرشی است. در شبکه‌ی یوی سوک اتصالات پرشی با اتصالات عمیق جایگزین می‌شوند. شکل ۵-۱ به تفاوت این شبکه با سایر شبکه‌های پیش‌تر تعریف شده می‌پردازد. در شبکه‌ی یوی سوک علاوه بر اتصالات پرشی کدگذار و کدگشای متناظر در هر سطح، اتصالاتی که پیش از این در شبکه‌ی یو وجود داشت، اتصالات پرشی از سطوح عمیق‌تر کدگذار نیز به هر سطح کدگشا متصل می‌شود. این اتصالات عمیق سبب می‌شوند که هم جزئیات ریز و هم کلیت معنایی در یک مجموعه ویژگی تمام مقیاس ظاهر شود.

شکل ۵-۲ روند ساخت یک مجموعه ویژگی تجمیع شده برای لایه‌ی دوم کدگشا ( $X_{De}^2$ ) را نشان می‌دهد. ابتدا همان طور که پیش‌تر هم گفته شد، یک اتصال پرشی از سطح متناظر در کدگذار ( $X_{En}^2$ ) بعد از عبور از یک کانولوشن  $3 \times 3$  در لایه‌ی مورد نظر دریافت می‌شود. و بعد اتصالات لایه‌های عمیق‌تر در کدگذار در این مثال لایه‌های ۳ و ۴ پس از عبور از یک کانولوشن  $3 \times 3$  و یک نمونه افزایشی به مجموعه ویژگی دریافت شده الحاق می‌شوند و در نهایت اتصال لایه‌ی قبلی در کدگشا نیز به این مجموعه ویژگی اضافه می‌شود و این مجموعه ویژگی با عبور از دو لایه‌ی کانولوشن  $3 \times 3$  و نرمال‌سازی دسته‌ای و تابع فعال‌سازی  $ReLU$  با یکدیگر ترکیب می‌شوند و مجموعه ویژگی نهایی را تولید می‌کنند. توضیحات گفته شده را می‌توان در فرمول زیر برای لایه‌ی  $i$  ام از کدگشا خلاصه کرد:



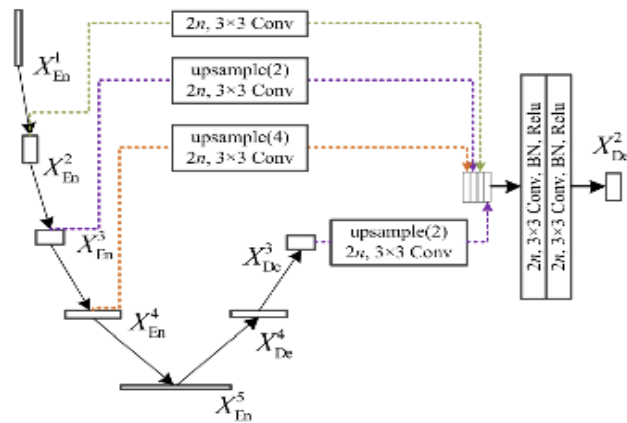
شکل ۵-۱: مقایسه‌ی اتصالات در چهار شبکه‌ی یو، یو++، یو+ و یوی سوک [۳]

$$X_{De}^i = \begin{cases} \Theta([\Gamma(X_{En}^i), \Gamma(\mathcal{U}(X_{En}^k))_{k=i+1}^{N-1}, \Gamma(\mathcal{U}(X_{De}^{i+1}))]) & i = 1, \dots, N-2 \\ \Theta([\Gamma(X_{En}^i), \Gamma(\mathcal{U}(X_{En}^{i+1}))]) & i = N-1 \end{cases} \quad (1-5)$$

که  $\mathcal{H}(\cdot)$  همان تابع ترکیب ویژگی‌هاست که همان طور که پیش‌تر هم گفته شد شامل دو لایه ی کانولوشن  $3 \times 3$  و نرمال‌سازی دسته‌ای و تابع فعال‌سازی است.  $\gamma(\cdot)$  عملگر کانولوشن را نشان می‌دهد و  $\mathcal{U}(\cdot)$  عملگر نمونه‌افزایی با استفاده از درون‌یابی خطی را نشان می‌دهد و  $[\cdot]$  عملگر الحاق را نشان می‌دهد.

## ۵-۲ آموزش شبکه

برای آموزش مناسب شبکه‌ی یوی سوک انتخاب مناسب تابع خطا اهمیت بسیاری دارد. در این مقاله برای انتخاب تابع خطا از یک رویکرد تجربی استفاده شده است. به این صورت که تعدادی از توابع خطا



شکل ۵-۲: چگونگی ساخت یک مجموعه ویژگی‌های جمع شده برای لایه‌ی دوم کدگشا [۳]

method	Dice
$U - net (Vgg - 16)$	0.92060
$U - net (ResNet - 101)$	0.93870
$U - net + + (Vgg - 16)$	0.92708
$U - net + + (ResNet - 101)$	0.94750
$U - net3+ (Vgg - 16)$	0.95500
$U - net3+ (ResNet - 101)$	0.96010
$ELU - net (Vgg - 16 \square DFK)$	0.96059
$ELU - net (ResNet34 + DFK)$	0.97365

جدول ۵-۱: نتایج به دست آمده از شبکه‌ی یوی سوک در مقایسه با سایر شبکه‌ها بر روی مجموعه داده‌ی  $ISBI$  [۳]

در نظر گرفته شده‌اند و یک یا ترکیبی از آن‌ها برای قطعه‌بندی چندین مجموعه داده به کار گرفته شده‌اند و سپس از معیار ضریب تشابه سرنسن-دایس که در فصل قبلی معرفی شد برای انتخاب بهترین تابع یا بهترین ترکیب توابع استفاده شده است. همچنین آموزش شبکه با بهینه‌ساز آدام انجام شده است.

### ۳-۵ نتایج

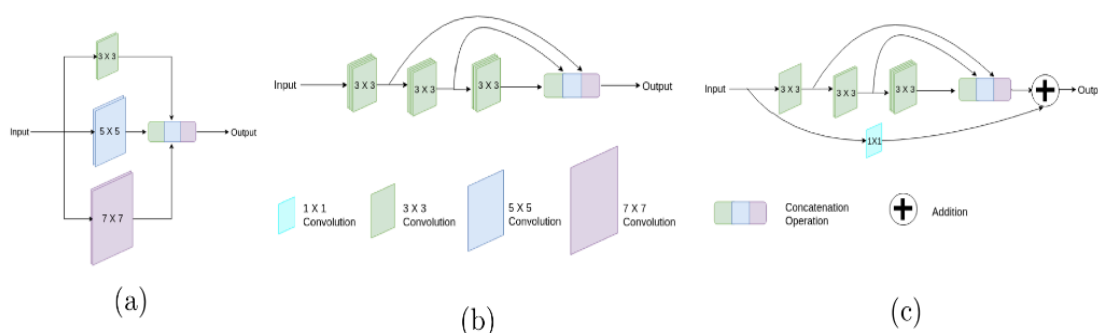
ارزیابی شبکه بر روی چندین مجموعه داده‌ی تصاویر پزشکی انجام شده است. در اینجا برای مقایسه صرفاً یک نمونه از این نتایج آورده شده است. جدول ۵-۱ نتایج به دست آمده از شبکه‌ی یوی سوک را در مقایسه با شبکه‌های پیش‌تر معرفی شده روی مجموعه داده‌ی  $LiTs ISBI$  2017 نشان می‌دهد. همان‌طور که مشاهده می‌کنید این روش نسبت به سایر روش‌های پیش‌تر معرفی شده ضریب تشابه بیشتری دارد.

## فصل ششم

### شبکه‌ی یوی چند باقی‌مانده‌ای

در فصل مقدمه ذکر شد که شبکه‌ی یو دو مشکل اساسی دارد. مشکل اول عدم مقاومت آن در برابر تغییر مقیاس و مشکل دوم وجود فاصله‌ی معنایی بین دو طرف یک اتصال پرشی بود. در فصل‌های گذشته نشان داده شد که چگونه ایده‌های مختلفی برای حل هر کدام از این مشکلات ارائه شد و عمده‌ی این ایده‌ها از طریق تغییر مسیرهای پرشی و افزایش اتصالات بین کدگذار و کدگشا اعمال می‌شد. شبکه‌ی یوی چند باقی‌مانده‌ای با این هدف توسعه داده شد که هر دو مشکل را با رویکردی متفاوت حل کند.

یکی از رویکردهای موثر در شبکه‌های کانولوشنی برای حل مساله‌ی تنوع مقیاس توسط زگدی و همکاران [۱۰] در معرفی بلوک دریافت<sup>۱</sup> مطرح شد. ایده‌ی اصلی این بود که استفاده‌ی موازی از فیلترهای کانولوشن با اندازه‌های مختلف موجب کشف ویژگی‌ها در مقیاس‌های مختلف می‌گردد. مشابه همین ایده برای غنی تر کردن یک شبکه‌ی یو و توانمند سازی آن در تحلیل چندمقیاسی، می‌توانیم از کانولوشن‌های با اندازه‌ی هسته‌ی متفاوت به صورت موازی استفاده کنیم. شکل ۱-۶ قسمت  $a$  نمونه‌ای از این بلوک را نشان می‌دهد.



شکل ۱-۶: ساخت مرحله به مرحله‌ی بلوک چند باقی‌مانده‌ای [۶]

بنابراین ایده‌ی اولیه‌ی شبکه‌ی یوی چند باقی‌مانده‌ای تعویض لایه‌های کانولوشن موجود در ساختار شبکه‌ی یو با بلوک‌های شبیه به بلوک دریافت بود تا بتواند به کمک این تعویض ویژگی‌های تصویر را در مقیاس‌های مختلفی یاد بگیرد. این موضوع علیرغم بهبود عملکرد شبکه‌ی یو به دلیل حجم زیاد محاسبات و تعداد پارامترهای زیاد در عمل قابل بهره‌برداری نبوده و باعث عدم کارایی شبکه‌ی یوی باقی‌مانده‌ای از نظر حجم محاسبات می‌گردد.

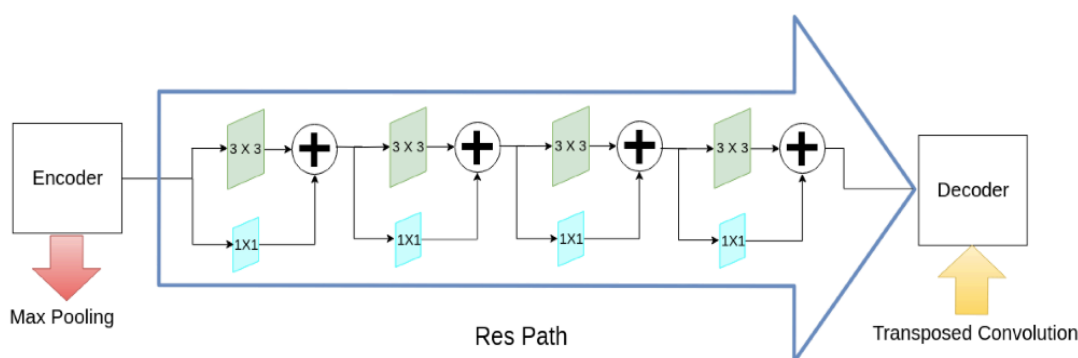
برای کاهش این حجم محاسباتی می‌توانیم کانولوشن‌های با هسته‌ی بزرگتر را با کانولوشن‌های متوالی کوچک‌تر و سبک‌تر جایگزین کنیم شکل ۱-۶ قسمت  $b$  جابه‌جایی فیلترهای  $7 \times 7$  و  $5 \times 5$  را با فیلترهای  $3 \times 3$  نشان می‌دهد و خروجی‌های کانولوشن‌های  $3 \times 3$  ی دوم و سوم تخمین نسبتاً مناسبی از اعمال کانولوشن‌های  $5 \times 5$  و  $7 \times 7$  می‌دهند. در نهایت از الحاق خروجی هر سه مرحله کانولوشن  $3 \times 3$  ویژگی‌های

<sup>1</sup>inception block



فضایی در مقیاس‌های مختلف ساخته می‌شود. در عمل مشاهده شده است که عملکرد این بلوک سبک مشابه همان بلوک دریافت تعریف شده در قسمت  $a$  شکل ۱-۶ می‌باشد. همچنین در عمل نشان داده شده است که اضافه کردن یک اتصال باقی‌مانده‌ای<sup>۲</sup> در قطعه‌بندی تصاویر پزشکی موثر است [۴] و از این رو در این بلوک یک اتصال باقی‌مانده‌ای نیز اضافه شده است. شکل ۱-۶ قسمت  $c$  بلوک چند باقی‌مانده‌ای نهایی را نشان می‌دهد.

مشکل دوم در یک شبکه‌ی یو فاصله‌ی معنایی میان دو طرف یک اتصال پرشی کدگذار و کدگشا بود. مشکلی که به طور خاص در شبکه‌ی یو++ به دنبال رفع آن بودیم. در شبکه‌ی یوی چندباقی‌مانده‌ای نیز مانند شبکه‌ی یو++ از کانولوشن در اتصالات پرشی برای حل این مشکل بهره گرفته شده است. در کنار این اتصالات پرشی حاوی کانولوشن یک اتصال باقی‌مانده‌ای نیز اضافه شد که باعث یادگیری آسان‌تر شبکه می‌گردد. فرضیه‌ی اصلی در این شبکه این است که اضافه کردن این پردازش کانولوشنی در اتصالات پرشی موجب می‌شود تا اطلاعات ورودی از کدگذار به سطح معنایی کدگشا برسند و دو طرف اتصال پرشی از نظر معنایی هم سطح گردند. شکل ۲-۶ فرآیند انتقال ویژگی از کدگذار به کدگشا را در یک اتصال پرشی نشان می‌دهد که شامل کانولوشن‌های متوالی و اتصالات باقی‌مانده‌ای می‌باشد.



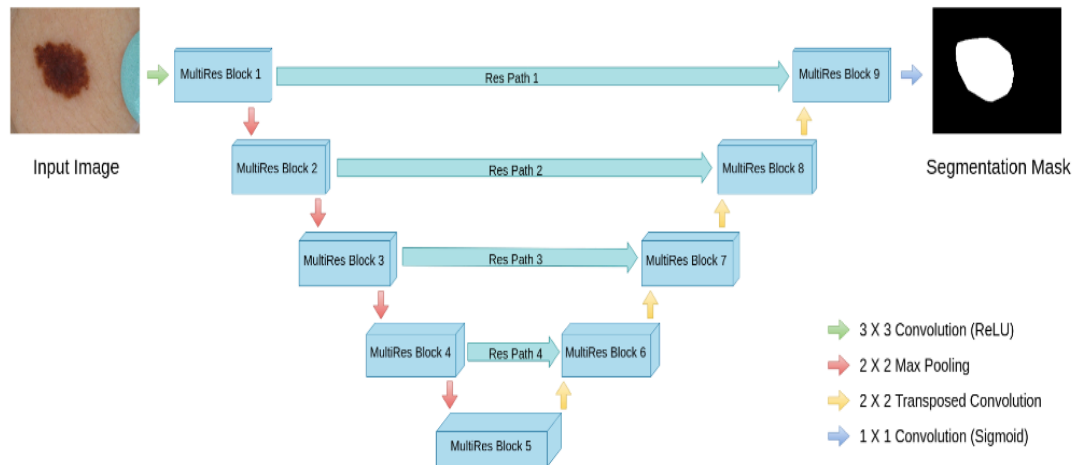
شکل ۲-۶: مسیر باقی‌مانده‌ای در شبکه‌ی یوی چند باقی‌مانده‌ای [۶]

## ۱-۶ معماری شبکه

در این ساختار به جای دو بلوک متوالی کانولوشن در ساختار اصلی شبکه‌ی یو یک بلوک چند باقی‌مانده‌ای قرار می‌گیرد که ساختار این بلوک در قسمت  $c$  شکل ۱-۶ قابل مشاهده است. همچنین به جای اتصالات پرشی از مسیرهای باقی‌مانده‌ای که در شکل ۲-۶ قابل مشاهده است، استفاده شده است. با توجه به اینکه هر چقدر به سمت عمق کدگذار حرکت می‌کنیم فاصله‌ی معنایی میان ویژگی‌های کدگذار و کدگشا کمتر می‌شود؛ می‌توانیم با حرکت به سمت عمق کدگذار از تعداد لایه‌های کانولوشن کم کنیم و به همین

<sup>۲</sup>residual connection

ترتیب در شبکه‌ی چند باقی‌مانده‌ای از پایین به بالا به ترتیب ۱، ۲، ۳، ۴ بلوک کانولوشن داریم. همه‌ی لایه‌های کانولوشن به کار گرفته شده در شبکه به جز لایه‌ی خروجی از تابع فعال‌سازی  $ReLU$  استفاده می‌کنند و لایه‌ی خروجی نیز مشابه شبکه‌ی یو از تابع فعال‌سازی سیگموید استفاده می‌کند. معماری کامل این شبکه در شکل ۳-۶ آورده شده است.



شکل ۳-۶: معماری شبکه‌ی یوی چند باقی‌مانده‌ای [۶]

## ۲-۶ آموزش شبکه

در این شبکه هدف پیش‌بینی تعلق هر پیکسل به ناحیه‌ی مورد نظر (مثلاً تومور یا جراحی) یا پس زمینه می‌باشد. بنابراین می‌توان تابع خطا را به صورت هر تابع خطای مساله‌ی دسته‌بندی دو کلاسه تعریف کرد. در مقاله این تابع همان تابع آنترپی متقابل دودویی است که به صورت زیر تعریف می‌شود:

$$CrossEntropy(X, Y, \hat{Y}) = \sum_{px \in X} -(y_{px} \log(\hat{y}_{px}) + (1 - y_{px}) \log(1 - \hat{y}_{px})) \quad (۱-۶)$$

که  $X$  تصویر اصلی و  $Y$  ماسک قطعه بندی حقیقی و  $\hat{Y}$  ماسک پیش بینی شده است. برای یک پیکسل  $p_x$  شبکه  $\hat{Y}_{p_x}$  را پیش بینی میکند و برچسب حقیق آن هم  $Y_{p_x}$  است. برای یک دسته شامل  $n$  تصویر تابع خطا به صورت زیر می‌باشد:

<i>model</i>	<i>parameters</i>
<i>U – net</i>	7,759,521
<i>MultiResU – net</i>	7,262,750

جدول ۶-۱: مقایسه‌ی تعداد پارامترهای شبکه‌ی یو و شبکه‌ی یوی چند باقی‌مانده‌ای [۶]

<i>Modality</i>	<i>MultiResUNet</i>	<i>U – Net</i>	<i>Relative Improvement</i>
<i>Fluorescence Microscopy</i>	$80.2988 \pm 0.3717$	$76.4277 \pm 4.5183$	5.065
<i>Electron Microscopy</i>	$82.0574 \pm 1.5953$	$74.4984 \pm 1.4704$	10.1465
<i>Dermoscopy</i>	$91.6537 \pm 0.9563$	$89.3027 \pm 2.1950$	2.6326
<i>Endoscopy</i>	$87.9477 \pm 0.7741$	$87.4092 \pm 0.7071$	0.6161
<i>MRI</i>	$78.1936 \pm 0.7868$	$77.1061 \pm 0.7768$	1.4104

جدول ۶-۲: جدول نتایج ارزیابی شبکه‌ی یو چند باقی‌مانده‌ای روی پنج مجموعه داده‌ی تصاویر پزشکی [۶]

$$j = \frac{1}{n} \sum_{i=1}^n \text{CrossEntropy}(X_i, Y_i, \hat{Y}_i) \quad (۲-۶)$$

هدف شبکه کمینه کردن معیار آنتروپی متقابل دودویی است که با استفاده از بهینه‌ساز آدام صورت می‌گیرد. شبکه در این مقاله برای ۱۵۰ گام آموزش داده شده است و علت انتخاب این مقدار گام هم این است که بعد از این تعداد دیگر بهبودی حاصل نمی‌شده است.

### ۳-۶ نتایج

ارزیابی این شبکه بر روی پنج مجموعه داده‌ی تصاویر پزشکی انجام شده است و مدل پایه برای مقایسه شبکه‌ی یو بوده است. جدول ۶-۱ مقایسه‌ی تعداد پارامترهای این دو مدل را نشان می‌دهد که همان‌طور که مشاهده می‌کنید تعداد پارامترهای شبکه‌ی یوی چند باقی‌مانده‌ای اندکی کمتر از شبکه‌ی یو می‌باشد با این حال نتایج آن بر روی مجموعه داده در مقایسه با شبکه‌ی یو بهبود یافته است. این نتایج در جدول ۶-۲ آورده شده است.

## کتاب نامه

- [1] Badrinarayanan, Vijay, Kendall, Alex, and Cipolla, Roberto. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.
- [2] Ciresan, Dan, Giusti, Alessandro, Gambardella, Luca, and Schmidhuber, Jürgen. Deep neural networks segment neuronal membranes in electron microscopy images. *Advances in neural information processing systems*, 25, 2012.
- [3] Deng, Yunjiao, Hou, Yulei, Yan, Jiangtao, and Zeng, Daxing. Elu-net: an efficient and lightweight u-net for medical image segmentation. *IEEE Access*, 10:35932–35941, 2022.
- [4] Drozdal, Michal, Vorontsov, Eugene, Chartrand, Gabriel, Kadoury, Samuel, and Pal, Chris. The importance of skip connections in biomedical image segmentation. In *International Workshop on Deep Learning in Medical Image Analysis, International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, pages 179–187. Springer, 2016.
- [5] Huang, Huimin, Lin, Lanfen, Tong, Ruofeng, Hu, Hongjie, Zhang, Qiaowei, Iwamoto, Yutaro, Han, Xianhua, Chen, Yen-Wei, and Wu, Jian. Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1055–1059. IEEE, 2020.

- [6] Ibtehaz, Nabil and Rahman, M Sohel. Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural networks*, 121:74–87, 2020.
- [7] Lee, Chen-Yu, Xie, Saining, Gallagher, Patrick, Zhang, Zhengyou, and Tu, Zhuowen. Deeply-supervised nets. In *Artificial intelligence and statistics*, pages 562–570. PMLR, 2015.
- [8] Long, Jonathan, Shelhamer, Evan, and Darrell, Trevor. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [9] Ronneberger, Olaf, Fischer, Philipp, and Brox, Thomas. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.
- [10] Szegedy, Christian, Liu, Wei, Jia, Yangqing, Sermanet, Pierre, Reed, Scott, Anguelov, Dragomir, Erhan, Dumitru, Vanhoucke, Vincent, and Rabinovich, Andrew. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [11] Zhou, Zongwei, Rahman Siddiquee, Md Mahfuzur, Tajbakhsh, Nima, and Liang, Jianming. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer, 2018.