نام و نام خانوادگی	**علی رنجبری – امیرحسین علیزاده**
شماره دانشجویی	**۸۱۰۱۹۸۵۷۰ - ۸۱۰۹۱۷۵۴۶**
تاریخ ارسال گزارش	**۱۴۰۱.۰۹.۱۵**

به نام خدا
دانشگاه تهران
دانشکده مهندسی برق و کامپیوتر

درس شبکه‌های عصبی و یادگیری عمیق
تمرین سوم

فهرست

شکل‌ها

جدول‌ها

پاسخ 1. آشنایی با یادگیری انتقالی Transfer Learning

۱) مقاله مربوطه مقاله ResNet بود که باید با استفاده از این شبکه به تشخیص سرطان روده پرداخته بود . دیتاستی که در این مقاله روی آن کار شده بود شامل تصاویری از غده در روده‌ی بزرگ است. که در کل شامل ۱۶۵ عکس است که ۷۴ عدد از آنها غدد خوش خیم و ۹۱ عکس دیگر بدخیم هستند.

در ابتدا برای پیش‌پردازش عکس ها ابتدا عکس ها grayscale شدند و سپس به خاطر کنتراست پایین با استفاده از تکنیک CLAHE کنتراست تصاویر را یکنواخت تر کردند تا برای دسته بندی در شبکه مورد نظر آماده شوند. (تصوری ۱)

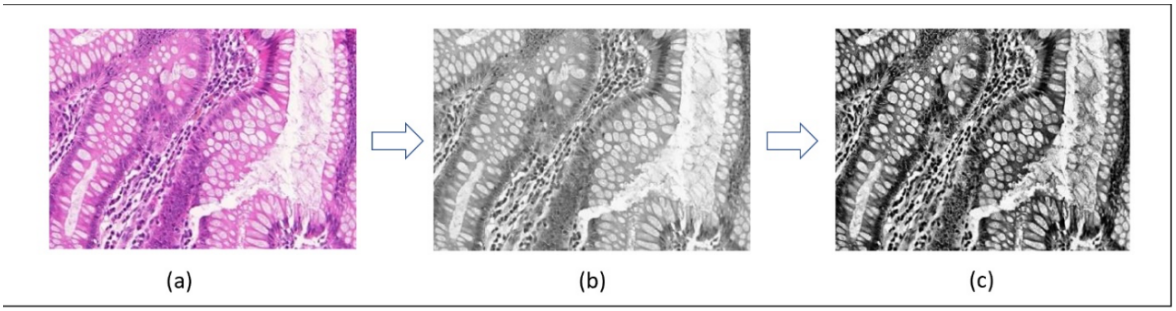


Fig. 4. The pre-processing process: a) input dataset; b) grayscale image; (c) and uniformity of contrast performed with CLAHE

بعد از پیش پردازش سپس با استفاده از ResNet-18 یا ResNet-50 به دسته بندی داده ها پرداخته شد (Optimizer استفاده شده در اینجا SGD با استفاده از momentum است) و با کمک سه شاخص Accuracy, Sensitivity, Specificity کارکرد شبکه ارزیابی شد. این سه شاخص را میتوانیم در زیر ببینیم:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

شمای کلی کار ها هم در شکل ۲ قابل مشاهده است.

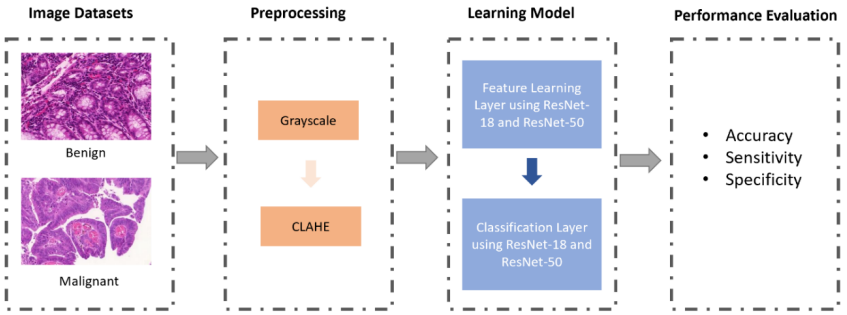


Fig. 2. Research Flow for Proposed Method

برای آموزش و ارزیابی در این مقاله داده ها را با سه روش متفاوت دسته بندی کردند در مدل اول ۶۰ درصد برای آموزش و ۴۰ درصد برای ارزیابی در مدل دوم ۷۵ درصد برای آموزش و ۲۵ درصد برای ارزیابی و در مدل

سوم ۸۰ درصد برای آموزش و فقط ۲۰ درصد برای ارزیابی در نظر گرفته شده بود. مدل ها با تابع خطای Cross-entropy Loss و SGD Optimizer آموزش داده شده اند. بیشترین accuracy مربوط به تقسیم بندی ۸۰ به ۲۰ بود که دقت ۸۵٪ گرفت و بیشترین sensitivity مربوط به شبکه ۷۵ و ۲۵ بود که دقت ۹۶٪ گرفت جدول ۱ به مقایسه این مدل های میپردازد.

Table 1. Comparison of Accuracy, Sensitivity, Specificity Value in Several Testing Datasets for ResNet-18

Training Data: Testing Data	***Accuracy***	***Sensitivity***	***Specificity***
60%:40%	73%	64%	83%
75%:25%	81%	96%/	63%
80%:20%	85%	83%	87%

۲) شبکه ResNet در چند سایز متفاوت وجود دارد که در جدول ۲ همه معماری ها به همراه لایه ها و مشخصات آنها قابل مشاهده است.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

جدول ۲. معماری شبکه های مختلف ResNet

پیش‌پردازش های لازم برای ResNet با استناد به سایت Pytorch ورودی شبکه باید دارای سه کانال رنگی به اندازه ی حداقل ۲۲۴ پیکسل برای هر محور است. همچنین عکس ها باید در بازه‌ی [0,1] باشند و با استفاده از داده هایی که در این سایت قابل مشاهده است normalize شوند.

معماری ResNet به دلیل داشتن Residual ها اطلاعات قبلی را هم حفظ میکنند و دوباره به شبکه تزریق می‌کنند به همین علت میتوانند عملکرد قابل قبولی از خود ارائه دهند.

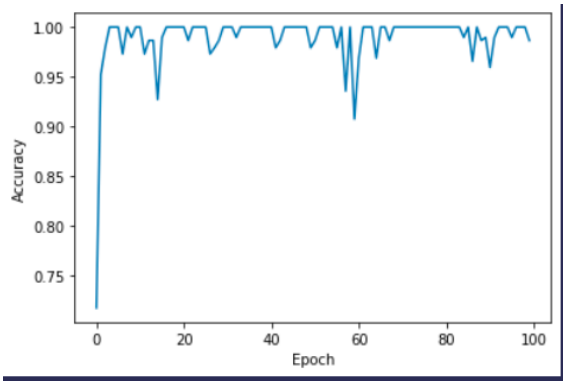
۳) شبکه ResNet رودی داده های ImageNet 2012 آموزش داده شده است که شامل ۱۰۰۰ کلاس مختلف داده و ۱.۲۸ میلیون تصویر برای آموزش است و ۵۰ هزار تصویر برای ارزیابی است. کلاس های مختلف که این شبکه میتوانند تشخیص دهد در این لینک قابل مشاهده است.

اما اگر داده هایی که داشتیم در این دسته بندی قرار نداشت میتوانیم با استفاده از تکنیک Transfer Learning از شبکه ResNet به عنوان Backbone استفاده کنیم و مثلا بعد از آخرین لایه آن یک لایه خروجی تعریف کنیم که به تعداد دلخواه ما دسته بندی داشته باشد و سپس شبکه را با داده های خودمان آموزش دهیم چون یک شبکه قوی که قبلا آموزش دیده است در اینجا وجود دارد آموزش زمان کمتری را میگیرد و به دقت های بهتری دست پیدا میکنیم.

۴) دیتای مربوطه که در این سایت در اختیار ما قرار داده شده بود در گوگل کولب لود شدند و روی آن کار شد که میتوانید در فایل های jupyter این تمرین مشاهده کنید.

همانطور که در بخش یک اشاره شد این دیتا شامل ۱۶۵ عکس متفاوت از غده در روده بزرگ است که دو دسته خوش‌خیم و بدخیم دارد که باید آموزش ببینند.

۵) شبکه ResNet با استفاده از pytorch لود شد و بر روی داده مورد نظر آموزش داده شد نمودار دقت آن در هر epoch به شکل (شکل ۳) مقابل است .



شکل ۳. نمودار پیشرفت دقت در هر epoch شبکه ResNet

در داده هایی که در اختیار داشتیم در دسته داده تست مختلف در اختیار ما قرار داده شده بود که ما در اینجا از ادغام داده های آموزش و دو دسته داده ارزیابی خودداری کردیم و طبق همان چیزی که تقسیم بندی شده بود آموزش و تست را انجام دادیم. داده ها شامل TestA و TestB بودند که عملکرد شبکه روی هر کدام از آنها در شکل ۴ قابل مشاهده است.

```
Accuracy: 0.939, Precision: 0.935, F1 Score: 0.933

Accuracy: 0.906, Precision: 0.786, F1 Score: 0.812
```

شکل ۴. قسمت بالا مربوط به TestA و قسمت پایین مربوط به TestB است.

پاسخ ۲ - آشنایی با تشخیص چهره مسدود شده

۱) در این مقاله از سه شبکه مختلف استفاده شده RSPNet و DeepLabv3+ و برای استخراج ویژگی ها از ResNet-101 به عنوان backbone استفاده شده است. همچنین از SegFormer و MIT-B5 به عنوان backbone استفاده شد. بعد از آموزش هم این شبکه ها روی دو دیتاست COFW و RealOcc-Wild تست شدند. همچنین عملکرد شبکه ها روی هر دیتاست در جدول ۳ قابل مشاهده است. همانطور که مشخص است در کل SegFormer عملکرد بهتری از خود نشان داده است.

	Quantity	**RealOcc** (mIoU)			**COFW (Train)** (mIoU)			**RealOcc-Wild** (mIoU)		
		PSPNet	DeepLabv3+	SegFormer	PSPNet	DeepLabv3+	SegFormer	PSPNet	DeepLabv3+	SegFormer
C-Original	29,200	89.52	88.13	88.33	89.64	88.62	91.36	85.21	82.05	85.24
C-CM	29,200	96.15	96.13	97.42	91.82	92.77	**94.87**	91.33	91.01	95.16
C-WO	24,602	89.38	89.01	91.36	89.53	88.97	92.24	83.86	84.14	86.72
C-WO + C-WO-NatOcc	24,602 + 49,204	96.65	96.51	97.30	90.71	91.21	94.30	91.34	91.70	94.17
C-WO + C-WO-NatOcc-SOT	24,602 + 49,204	96.35	96.59	97.18	**92.32**	91.74	93.55	**93.26**	92.69	94.27
C-WO + C-WO-RandOcc	24,602 + 49,204	95.09	95.21	96.53	90.82	91.35	93.14	89.54	89.68	92.84
C-WO + C-WO-Mix	24,602 + 73,806	96.55	96.66	97.37	90.99	91.20	93.74	92.14	91.84	94.40
C-CM + C-WO-NatOcc	29,200 + 49,204	**97.28**	**97.33**	97.95	91.61	92.66	94.86	92.13	**93.81**	**95.43**
C-CM + C-WO-NatOcc-SOT	29,200 + 49,204	97.17	97.29	**98.02**	92.07	**92.91**	94.60	92.84	93.73	94.53

جدول ۳. ارزیابی کلی شبکه های مختلف

۲) دو Occluder مختلف در این مقاله استفاده شده NatOcc و RandOcc و دیتاستی که برای آموزش این شبکه ها استفاده شده بودند دیتاست CelebAMask-HQ-WO بود و دقت هر کدام دیتاست با occluder های مختلف در جدول ۳ قابل مشاهده است همچنین تعریف هر کدام از دیتاست ها هم در جدول ۴ قرار دارد.

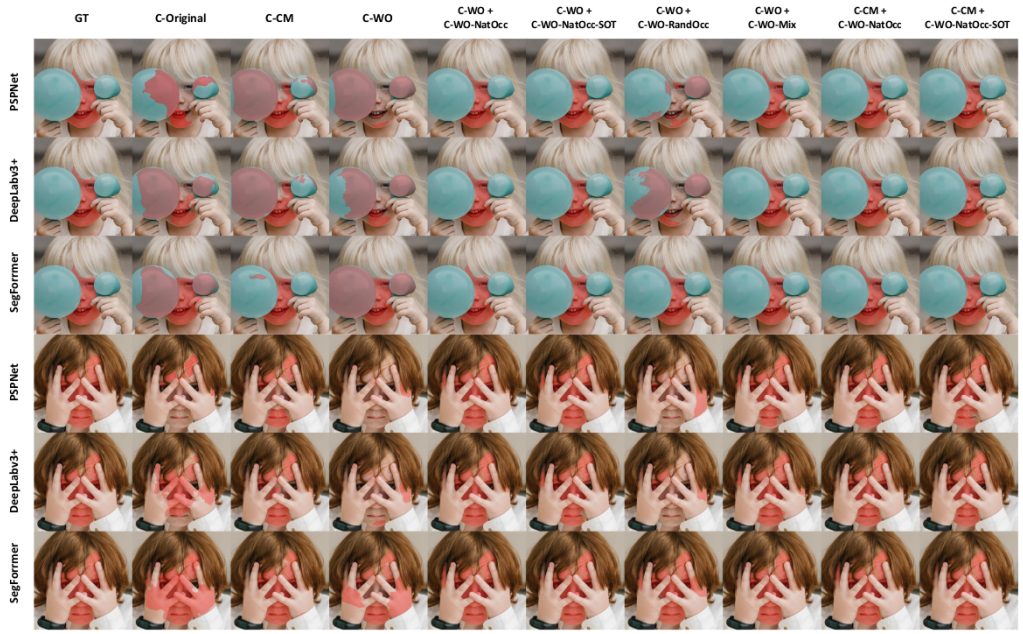
Class	Definition
C-Original	CelebAMask-HQ-WO (Train) and CelebAMask-HQ-O with original masks.
C-CM	CelebAMask-HQ-WO (Train) and CelebAMask-HQ-O with corrected masks.
C-WO	CelebAMask-HQ-WO (Train).
C-WO-NatOcc	One set of hand-occluded (without color transfer) face dataset and one set of COCO-occluded face dataset generated by NatOcc with C-WO.
C-WO-NatOcc-SOT	One set of hand-occluded (with color transfer) face dataset and one set of COCO-occluded face dataset generated by NatOcc with C-WO.
C-WO-RandOcc	Two sets of occluded face dataset generated by RandOcc with C-WO.
C-WO-Mix	Half set of C-WO-RandOcc and one set of C-WO-NatOcc.

جدول ۴. تعریف دیتاست های مختلف

همانطور که در جدول ۳ مشخص است دیتاست هایی که با استفاده از NatOcc مسدود شده بودند و آموزش داده شده بودند دقت در کل دقت بهتری دریافت کردند . البته همه این دیتاست ها از ترکیب دیتاست ها ساخته شده تا به دقت مورد نظر رسیده اما با مشاهده دقت ها میتوان پی برد که دقت کلی با استفاده از NatOcc بهتر از RandOcc است . اما ترکیب این دو مسدود کننده می تواند عملکرد بهتری هم داشته باشد.

۳) کلاس بندی کردن داده ها باعث عملکرد بهتر شبکه در تشخیص چهره مسدود شده میباشد. و همانطور که در جدول ۳ و شکل ۵ مشخص است عملکرد شبکه روی دیتاست C-WO که کلاس بندی نشده و فقط چهره های غیر مسدود شده دارد و همچنین C-Original که چهره مسدود شده دارد اما به صورت غیر دقیق annotate شدند بسیار ضعیف تر از دیتاست هایی است که دو کلاس چهره مسدود شده و غیر مسدود شده دارند و یا به صورت دستی یا با

استفاده از الگوریتم هایی مثل NatOcc و RandOcc مسدود و annotate شده اند است. بنابراین اگر دو کلاس تصویر مسدود شده و مسدود نشده دقیق داشته باشیم شبکه عملکرد بسیار دقیق تری خواهد داد.



۴) همانطور که مقاله و جدول ۳ دیدیم عملکرد شبکه SegFormer در اکثر دیتاست ها بهتر است و همچنین دیتاست C-WO-NatOcc هم دست های مسدود کننده و صورت دارای intensity و رنگ های متفاوت هستند و در این دیتاست از color transformer استفاده نشده است ولی همچنان عملکرد شبکه SegFormer بهتر از دو شبکه دیگر است بنابراین این شبکه برای تصاویر توصیه میشود.

۵)

پاسخ ۳ – تشخیص بلادرنگ اشیاء