



به نام خدا  
دانشگاه تهران  
دانشکده مهندسی برق و کامپیوتر



## درس شبکه‌های عصبی و یادگیری عمیق

### تمرین دوم

محمد ناصری	نام دستیار طراح	پرسش‌های ۱ و ۲
Mohammad.na3ri@gmail.com	رایانامه	
فاطمه جوادی یگانه	نام دستیار طراح	پرسش‌های ۳
Fatemeh.javadi.y@gmail.com	رایانامه	
۱۴۰۲.۰۹.۰۲	مهلت ارسال پاسخ	

## فهرست

قوانین	۱
پرسش ۱. تجزیه و تحلیل احساسات صورت مبتنی بر CNN	۱
۱-۱. معرفی مقاله	۱
۲-۱. پیش‌پردازش تصاویر و Data Augmentation	۱
۳-۱. پیاده‌سازی مدل AlexNet	۲
پرسش ۲ - پیاده‌سازی مدل VGGNet	۳
۱-۲. مدل VGGNet	۳
۲-۲. مدل MobileNet	۴
پرسش ۳ - تشخیص بیماران مبتلا به کووید با استفاده از عکس ریه	۵
۱-۳. معرفی مقاله	۵
۲-۳. جمع‌آوری داده و پیش‌پردازش تصاویر (۳۰ نمره)	۵
۳-۳. آموزش شبکه (۴۰ نمره)	۵
۴-۳. ارزیابی شبکه (۳۰ نمره)	۵
۵-۳. ارزیابی شبکه (امتیازی، ۱۰ نمره)	۶

## جدول‌ها

- جدول ۱ - معماری AlexNet ..... ۲
- جدول ۲ - معماری مدل VGGNet ..... ۳
- جدول ۳ - معماری MobileNet ..... ۴

قبل از پاسخ دادن به پرسش‌ها، موارد زیر را با دقت مطالعه نمایید:

- از پاسخ‌های خود یک گزارش در قالبی که در صفحه‌ی درس در سامانه‌ی Elearn با نام **REPORTS\_TEMPLATE.docx** قرار داده شده تهیه نمایید.
- پیشنهاد می‌شود تمرین‌ها را در قالب گروه‌های دو نفره انجام دهید. (بیش از دو نفر مجاز نیست و تحویل تک نفره نیز نمره‌ی اضافی ندارد) توجه نمایید الزامی در یکسان ماندن اعضای گروه تا انتهای ترم وجود ندارد. (یعنی، می‌توانید تمرین اول را با شخص A و تمرین دوم را با شخص B و ... انجام دهید)
- **کیفیت گزارش شما در فرآیند تصحیح از اهمیت ویژه‌ای برخوردار است؛** بنابراین، لطفاً تمامی نکات و فرض‌هایی را که در پیاده‌سازی‌ها و محاسبات خود در نظر می‌گیرید در گزارش ذکر کنید.
- در گزارش خود مطابق با آنچه در قالب نمونه قرار داده شده، برای شکل‌ها زیرنویس و برای جدول‌ها بالانویس در نظر بگیرید.
- الزامی به ارائه توضیح جزئیات کد در گزارش نیست، اما باید نتایج بدست آمده از آن را گزارش و تحلیل کنید.
- **تحلیل نتایج الزامی می‌باشد، حتی اگر در صورت پرسش اشاره‌ای به آن نشده باشد.**
- **دستیاران آموزشی ملزم به اجرا کردن کدهای شما نیستند؛** بنابراین، هرگونه نتیجه و یا تحلیلی که در صورت پرسش از شما خواسته شده را به طور واضح و کامل در گزارش بیاورید. در صورت عدم رعایت این مورد، بدیهی است که از نمره تمرین کسر می‌شود.
- **کدها حتماً باید در قالب نوت‌بوک با پسوند ipynb تهیه شوند، در پایان کار، تمامی کد اجرا شود و خروجی هر سلول حتماً در این فایل ارسالی شما ذخیره شده باشد.** بنابراین برای مثال اگر خروجی سلولی یک نمودار است که در گزارش آورده‌اید، این نمودار باید هم در گزارش هم در نوت‌بوک کدها وجود داشته باشد.
- **در صورت مشاهده‌ی تقلب امتیاز تمامی افراد شرکت‌کننده در آن، 100- لحاظ می‌شود.**
- تنها زبان برنامه نویسی مجاز **Python** است.
- **استفاده از کدهای آماده برای تمرین‌ها به هیچ وجه مجاز نیست.**

- نحوه محاسبه تاخیر به این شکل است: پس از پایان رسیدن مهلت ارسال گزارش، حداکثر تا یک هفته امکان ارسال با تاخیر (به ازای هر روز 5 درصد کسر نمره) وجود دارد، پس از این یک هفته نمره آن تکلیف برای شما صفر خواهد شد.
  - لطفاً گزارش، کدها و سایر ضمایم را به در یک پوشه با نام زیر قرار داده و آن را فشرده سازید، سپس در سامانه‌ی Elearn بارگذاری نمایید:
- HW [Number] \_ [Lastname] \_ [StudentNumber] \_ [Lastname] \_ [StudentNumber].zip
- (مثال: HW1\_Ahmadi\_810199101\_Bagheri\_810199102.zip)
- برای گروه‌های دو نفره، بارگذاری تمرین از جانب یکی از اعضا کافی است ولی پیشنهاد می‌شود هر دو نفر بارگذاری نمایند.

## پرسش ۱. تجزیه و تحلیل احساسات صورت مبتنی بر CNN

در این پرسش و پرسش دوم این تمرین قصد داریم تا با استفاده از شبکه CNN به تحلیل و طبقه‌بندی احساسات چهره از روی تصویر بپردازیم.

### ۱-۱. معرفی مقاله

مقاله<sup>۱</sup> مورد استفاده در این تمرین بر روی طراحی، استقرار و ارزیابی معماری‌های شبکه عصبی کانولوشنال (CNN) برای تجزیه و تحلیل احساسات چهره در دستگاه‌های تلفن همراه تمرکز دارد. برخلاف رویکردهای سنتی CNN، مدل‌های مستقر در دستگاه‌های تلفن همراه باید نیازهای ذخیره‌سازی را به حداقل برسانند و در عین حال عملکرد بالا را حفظ کنند. بنابراین برای دستیابی به این امر ۳ معماری متفاوت CNN در این مقاله ارائه شده است. نتایج به‌دست‌آمده نشان می‌دهد که معماری‌های پیشنهادی عملکرد مشابهی را نسبت به آخرین مدل‌های پیشرو در این زمینه دارند و این در حالی است که نیازهای ذخیره‌سازی را به حداقل می‌رسانند. معماری‌های مورد استفاده در این مقاله عبارتند از AlexNet، VGGNet و MobileNet

### ۲-۱. پیش‌پردازش تصاویر و Data Augmentation

(۱۰ نمره)

به منظور تولید تصاویر مناسب برای ورودی به مدل‌ها، تصاویر چهره با اندازه  $128 \times 128$  پیکسل استفاده شده است. این تصاویر با نمونه‌گیری از مجموعه‌داده AffectNet به‌دست‌آمده است. مجموعه اصلی شامل ۴۵۰ هزار تصویر چهره می‌باشد که به ۸ کلاس تقسیم شده‌اند. در این تمرین ما از ۱۰۰۰ و ۲۰۰ تصویر به ترتیب برای آموزش و Fine-tune کردن مدل استفاده خواهیم کرد. ([Link](#))

برای افزایش سرعت آموزش تمام داده‌های تصویر از [۰، ۲۵۵] به [۰، ۱] نرمال می‌شوند. همچنین جهت افزایش تعمیم‌پذیری مدل لازم است به مجموعه‌داده تصاویری با حالت‌های مختلف به صورت تصادفی اضافه کنیم. لازم است اضافه کردن داده برای مجموعه آموزشی با حالت‌های زیر را انجام بگیرد:

- چرخش تصاویر تا ۲۰ درجه
- Translation تا ۱۰٪ (در هر دو جهت x و y)
- چرخش در جهت x

---

<sup>۱</sup> [CNN-based Facial Affect Analysis on Mobile Devices](#)

در آخر نیاز است تا مجموعه داده به مجموعه های آموزش و ارزیابی تقسیم شوند.

## ۳-۱ پیاده سازی مدل AlexNet

(۹۰ نمره)

این معماری شامل مجموعه ای از هسته های کانولوشن با ابعاد کرنل رو به کاهش می باشد که از  $9 \times 9$  شروع می شود و به  $3 \times 3$  کاهش می یابد. جهت منظم سازی و آموزش سریع تر، هر بلوک کانولوشن از یک لایه کانولوشن دو بعدی معمولی و به دنبال آن یک لایه Batch-Normalization و یک لایه فعال سازی ReLU ساخته شده است. معماری همچنین کم عمق تر و باریک تر از طرح اصلی AlexNet است تا اندازه مدل را به حداقل برساند.

جدول ۱ - معماری AlexNet

Type	Shape	Output
Conv	$9 \times 9 \times 16$	$128 \times 128 \times 16$
MaxPool	$2 \times 2$	$64 \times 64 \times 16$
Conv	$7 \times 7 \times 32$	$64 \times 64 \times 32$
MaxPool	$2 \times 2$	$32 \times 32 \times 32$
Conv	$5 \times 5 \times 64$	$32 \times 32 \times 64$
MaxPool	$2 \times 2$	$16 \times 16 \times 64$
Conv	$3 \times 3 \times 128$	$16 \times 16 \times 128$
MaxPool	$2 \times 2$	$8 \times 8 \times 128$
Conv	$3 \times 3 \times 128$	$8 \times 8 \times 128$
MaxPool	$2 \times 2$	$4 \times 4 \times 128$
Flatten	2048	—
2xDense	1024	—
Dense	8 or 2	1 label or 2 floats

- ابتدا مدل را بر روی مجموعه Train آموزش داده و سپس Fine Tune کنید. (۵۰ نمره)
- نمودارهای Loss و Accuracy را رسم کرده و تحلیل کنید. (۱۰ نمره)
- نمودار ROC برای هر کلاس را رسم کرده و تحلیل کنید. (۱۰ نمره)
- مقادیر Precision، Recall و F1 را برای هر کلاس محاسبه و گزارش کنید. (۱۰ نمره)
- ماتریس Confusion را برای کلاس ها رسم کرده و تحلیل کنید. (۱۰ نمره)

## پرسش ۲ - پیاده‌سازی مدل VGGNet

این پرسش در ادامه پرسش اول مطرح شده است.

### ۲-۱. مدل VGGNet

(۱۰۰ نمره)

این معماری تقریباً شبیه به طراحی الهام گرفته شده از AlexNet در بالا است

جدول ۲ - معماری مدل VGGNet

Type	Shape	Output
2xConv	$3 \times 3 \times 16$	$128 \times 128 \times 16$
MaxPool	$2 \times 2$	$64 \times 64 \times 16$
2xConv	$3 \times 3 \times 32$	$64 \times 64 \times 32$
MaxPool	$2 \times 2$	$32 \times 32 \times 32$
2xConv	$3 \times 3 \times 64$	$32 \times 32 \times 64$
MaxPool	$2 \times 2$	$16 \times 16 \times 64$
2xConv	$3 \times 3 \times 128$	$16 \times 16 \times 128$
MaxPool	$2 \times 2$	$8 \times 8 \times 128$
2xConv	$3 \times 3 \times 128$	$8 \times 8 \times 128$
MaxPool	$2 \times 2$	$4 \times 4 \times 128$
Flatten	2048	—
2xDense	1024	—
Dense	8 or 2	1 label or 2 floats

- ابتدا مدل را بر روی مجموعه Train آموزش داده و سپس Fine Tune کنید. (۶۰ نمره)
- نمودارهای Loss و Accuracy را رسم کرده و تحلیل کنید. (۱۰ نمره)
- نمودار ROC برای هر کلاس را رسم کرده و تحلیل کنید. (۱۰ نمره)
- مقادیر Precision، Recall و F1 را برای هر کلاس محاسبه و گزارش کنید. (۱۰ نمره)
- مدل VGGNet و AlexNet را با یکدیگر مقایسه کرده و علل تفاوت نتایج را تحلیل و گزارش کنید. (۱۰ نمره)



## ۲-۲ مدل MobileNet

(۱۰ نمره امتیازی)

. این معماری از لایه‌های کانولوشن قابل تفکیک در عمق و به دنبال آن لایه‌های کانولوشن معمولی برای حفظ عملکرد بالا و در عین حال به حداقل رساندن پیچیدگی معماری استفاده می‌کند. این منجر به معماری‌های شبکه بسیار کوچک‌تر و قابل تنظیم و مناسب برای استقرار در دستگاه‌های تلفن همراه می‌شود. (جهت گرفتن نمره این قسمت دقت بالای ۴۰ درصد مدنظر می‌باشد)

جدول ۳ - معماری MobileNet

Type	Shape	Stride	Output
Conv	$3 \times 3 \times 32$	2	$64 \times 64 \times 32$
DConv	$3 \times 3 \times 64$	1	$64 \times 64 \times 64$
DConv	$3 \times 3 \times 128$	2	$32 \times 32 \times 128$
DConv	$3 \times 3 \times 128$	1	$32 \times 32 \times 128$
DConv	$3 \times 3 \times 256$	2	$16 \times 16 \times 256$
DConv	$3 \times 3 \times 256$	1	$16 \times 16 \times 256$
DConv	$3 \times 3 \times 512$	2	$8 \times 8 \times 512$
5xDConv	$3 \times 3 \times 512$	1	$8 \times 8 \times 512$
DConv	$3 \times 3 \times 1024$	2	$4 \times 4 \times 1024$
DConv	$3 \times 3 \times 1024$	1	$4 \times 4 \times 1024$
GlobalAvePool	1024	—	—
Dense	8 or 2	—	1 label or 2 floats

- ابتدا مدل را بر روی مجموعه Train آموزش داده و سپس Fine Tune کنید.
- نمودارهای Loss و Accuracy را رسم کرده و تحلیل کنید.
- تفاوت این مدل با دو مدل قبل در چیست؟ علل نتایج به‌دست آمده را تحلیل و گزارش کنید.

## پرسش ۳- تشخیص بیماران مبتلا به کووید با استفاده از عکس ریه

### ۳-۱. معرفی مقاله

در [مقاله](#) مورد استفاده در این تمرین، با استفاده از عکس x-ray از ریه بیماران مبتلا به کووید و افراد عادی یک شبکه عصبی کانولوشنال (CNN) طراحی شده است. با توجه به خطر آفرین بودن این بیماری برای جان افراد، آموزش یک شبکه با دقت بالا اهمیت زیادی دارد.

### ۳-۲. جمع آوری داده و پیش پردازش تصاویر (۳۰ نمره)

در بخش جمع آوری داده، نحوه ی جمع آوری داده توضیح داده شده است. به این منظور دیتاست اولیه توضیح داده شده جمع آوری شده است و از اینجا قابل دریافت است. بنابراین کار برای شما از بخش پیش پردازش داده ها شروع می شود.

برای پیش پردازش ابتدا با توجه به توضیحات مقاله، data augmentation را انجام دهید. چهار نوع data augmentation داریم. توصیه می شود این ۴ نوع را به صورت گام به گام پیاده سازی کنید تا از مشکل overfitting در شبکه خود جلوگیری کنید یعنی ابتدا نوع اول را بر دیتاست اعمال کرده و بعد شبکه را آموزش دهید و نتایج را تحلیل کنید سپس نوع دو را اعمال کرده و دوباره آموزش دهید. همچنین لازم است که normalization هم انجام داده شود.

### ۳-۳. آموزش شبکه (۴۰ نمره)

شبکه CNN را با توجه به توضیحات داده شده، پیاده سازی کنید و بر روی داده های آموزش، آموزش دهید. نمودار Loss و Accuracy را برای داده های آموزش و ولیدیشن رسم کنید.

### ۳-۴. ارزیابی شبکه (۳۰ نمره)

شبکه آموزش داده شده را بر روی داده های تست، ارزیابی کنید و معیارهای Precision، Accuracy، Sensitivity، Specificity و F1 score را گزارش کنید. ماتریس آشفتگی را رسم کنید. فراموش نکنید که در مقاله داده شده در مرحله ی پیش پردازش از نظر یک پزشک متخصص برای انتخاب داده ها استفاده شده است که ما امکان انجام این کار را نداریم بنابراین ممکن است شبکه شما با اختلاف نسبت به مقاله عمل کند

### ۳-۵. ارزیابی شبکه (امتیازی، ۱۰ نمره)

دقت شبکه را با توجه به ستون اول جدول ۶ برای داده های تست با تعداد مختلف لایه برای شبکه گزارش کنید. برای اینکار لازم است که هربار شبکه را آموزش داده و تست بگیرید.