



بسم الله الرحمن الرحيم

یادگیری عمیق دکتر عمادالدین فاطمی زاده

دانشکده برق

تمرین سری ۲

شبکه های CNN

مقدمه ای بر این تمرین:

تمرین های بخش **CNN** ها به دلیل نزدیکی زیادی که به مباحث بینایی ماشین دارد، سعی شده است بسیار نزدیک به این حوزه طرح شود. بار عملی تمرین ها نسبت به تئوری ها بیشتر بوده و سعی شده که شما در عمل با عملکرد این شبکه ها بر روی تصاویر آشنا شوید.

ابزار شما برای این تمرین ها **Tensorflow-1,2** می باشد اما برای تمرین **CIFAR10** می توانید از **Keras** استفاده کنید و برای سوال امتیازی هم استفاده از **Pytorch** موردی ندارد. سعی کنید پاسخ تمرین را بسیار تمیز و خوانا بنویسید و تمامی شکل ها و نمودارها را به تفصیل در گزارش خود بیاورید.

کد های شما باید قابلیت بازیابی داشته باشند و بدون هیچ مشکلی اجرا شوند.

مسئله ۱: Network Design (۱۲ نمره)

(الف) یک شبکه عصبی کانولوشنی دارای ۴ لایه کانولوشن 3×3 با اندازه گام ۱ بدون لایه **pooling** فرض کنید. هریک از لایه های نورون آخر این شبکه چه تعداد از پیکسل های ورودی را پوشش می دهد؟

(ب) لایه **Pooling** چگونه و با اضافه کردن چه خواصی به شبکه های کانولوشنی قدرت و تعمیم پذیری آن را افزایش می دهد؟

(ج) فرض کنید در قسمت^۲ از شبکه سایز تصویر ورودی 35×35 و تعداد کانال های آن برابر ۱۶ است، سایز فیلتر مورد استفاده در لایه کانولوشن 3×3 با طول گام ۱ و تعداد فیلترها ۳۲ عدد باشد. ابعاد خروجی و همچنین تعداد پارامترهای این لایه از شبکه را مشخص کنید. (با **Padding same**)

(د) شبکه زیر را با **Padding same** در نظر بگیرید:

$[227 \times 227 \times 3] \rightarrow \text{Input}$

$\text{Conv1} = (256 \times 5 \times 5)$

$$\text{Conv2} = (128 \times 5 \times 5)$$

$$\text{Max-Pooling} = (5 \times 5, \text{Stride } 2 \times 2)$$

$$\text{Fully Connected Layer} = (512)$$

$$\text{Fully Connected Layer} = (100)$$

سایز خروجی در هریک از لایه‌های کانولوشن و **Pooling** و تعداد پارامترها برای تمام لایه‌ها مشخص کنید. سپس مشکلات این شبکه را شرح دهید و تغییرات لازم را برای بهبود آن انجام دهید و سپس شبکه خود را از لحاظ تعداد پارامتر با شبکه ارائه شده مقایسه کنید.

ه) میدانیم که شبکه‌های عصبی تنها با دو لایه مخفی خاصیت general function approximator بودن خود را حفظ می‌نمایند. با این وجود، چرا گاهی بیش از دو لایه مخفی برای شبکه‌ها استفاده می‌شود؟

خ) تفاوت vanishing gradient, exploding gradient را شرح دهید.

مسئله ۲. **Transposed Convolution**: (۸ نمره)

کانولوشن بین ورودی **X** و فیلتر **W** به صورت زیر است:

$$W = \begin{bmatrix} w_{0,0} & w_{0,1} \\ w_{1,0} & w_{1,1} \end{bmatrix}, X = \begin{bmatrix} x_{0,0} & x_{0,1} & x_{0,2} \\ x_{1,0} & x_{1,1} & x_{1,2} \\ x_{2,0} & x_{2,1} & x_{2,2} \end{bmatrix}$$

می‌توان عملیات کانولوشن را به صورت ضرب ماتریسی نوشت که ورودی و خروجی را به صورت یک بردار و فیلتر را به صورت یک ماتریس در نظر گرفت. ورودی **X** را به صورت بردار زیر نمایش می‌دهیم:

$$X = [x_{0,0}, x_{0,1}, \dots, x_{2,2}]^T$$

الف) عملیات کانولوشن بالا با **s=1** را به صورت ضرب ماتریسی **Y=AX** بنویسید،

این عملیات را یک بار به ازای **zero-padding** تکرار کنید.

ب) در مورد **Masked Convolution** و کاربرد آن تحقیق کنید.

ج) separable convolution را توضیح دهید.

مسئله ۳: Object Detection (۲۵نمره) – اختیاری با نمره اضافه

در بین مدل‌های تشخیص اشیاء، معماری‌های یکپارچه YOLO از لحاظ سرعت و دقت به نتایج چشمگیری رسیدند. این الگوریتم‌ها می‌توانند در زمان کمتری در مرحله آموزش به همگرایی برسند و در مرحله استنتاج، اشیاء را با سرعت بالایی تشخیص داده و دسته‌بندی کنند.

الف) فرض کنید می‌خواهیم شبکه‌های YOLO-v1 و YOLO-v3 را روی داده‌هایی که دارای 300 کلاس هستند آموزش دهیم. لایه آخر هر کدام از این شبکه‌ها دارای چه عمقی (channel) باید باشند؟ (تعداد پنجره‌های پیشبینی (Bounding box) را برابر با مقادیر ذکر شده در مقاله‌های اصلی شبکه در نظر بگیرید)

ب) عملکرد YOLO-v1 روی داده‌هایی که دارای اشیایی با تداخل (overlap) بالا هستند، چه نقص عمده‌ای دارد؟ توضیح دهید. برای حل این مشکل چه روشی در YOLO-V2 اتخاذ شده است؟

ج) در الگوریتم YOLO-v3 برای تشخیص صحیح‌تر داده‌هایی که دارای اشیایی با پراکندگی اندازه‌های بالا (High variation size) هستند، چه روش‌هایی پیشنهاد شده است؟ توضیح دهید.

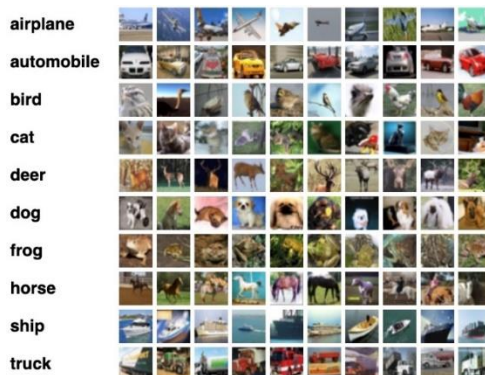
تمرین‌های عملی

CNN

۱ تلفیق MLP, CNN (۲۵ نمره)

در این تمرین می‌خواهیم به مقایسه بین شبکه‌هایی که در تمرین ۱ یاد گرفتید و تمرین دوم بپردازیم.

هدف در این تمرین ایجاد یک طبقه‌بند برای طبقه‌بندی مجموعه داده CIFAR-10 با استفاده از شبکه‌های MLP است. این مجموعه داده (نمونه‌ای از آن را تصویر زیر مشاهده می‌نمائید) شامل ۶۰ هزار تصویر رنگی است که در 10 کلاس دسته‌بندی شده است.



به طور معمول از 50 هزار تصویر آن به عنوان مجموعه داده آموزشی استفاده می‌کنند و 10 هزار تصویر را به عنوان مجموعه داده تست استفاده می‌کنند. شما این مجموعه داده را دانلود و 10 تصویر ابتدای آن را همراه با نام آن شی نمایش داده و سپس داده‌ها را به سه بخش آموزش، تست و ارزیابی تقسیم کنید.

در ادامه پیش پردازش‌های لازم را انجام دهید تا داده‌ها برای آموزش شبکه عصبی آماده بشوند. حال با توجه به آموخته‌های کلاس درس، یک شبکه MLP طراحی نمائید. هدف این سوال پیاده سازی شبکه عصبی، بررسی تاثیر تغییرات هایپرپارامترها و حل چالش مجموعه داده نامتوازن است.

فرضیات مسئله:

- تعداد لایه های مخفی را برابر 2 در نظر بگیرید
- از روش SGD استفاده نمایید.
- هایپرپارامترها مانند تابع خطا، نرخ یادگیری و ... را نیز خودتان انتخاب نمائید برای این کار می توانید از Grid Search یا آزمون و خطا استفاده نمائید. پارامترهای انتخاب شده را در گزارش بیاورید.
- موارد زیر را در گزارش برای قسمت های ب و ج بیاورید:
در دو نمودار جداگانه تغییرات دقت و خطای مدل در هر دور را برای داده ی ارزیابی و آموزش حالت های خواسته شده نشان دهید
- همچنین خطا، دقت و ماتریس آشفتگی (Confusion Matrix) را برای داده ی تست محاسبه کنید
- الف) لازم است برای حل این مسئله از روش Stochastic mini batch based استفاده شود. از سه دسته با اندازه های 64، 32 و 256 استفاده نمائید و تاثیر تفاوت اندازه دسته ها را در دقت و زمان آموزش شبکه بررسی نمائید.
- نکته: دقت نمائید در این آزمایش ها بقیه هایپرپارامترها ثابت هستند (توابع فعالساز همه لایه های مخفی را ثابت در نظر بگیرید).
- ب) مجموعاً 2 مرتبه توابع فعالساز را در لایه های مخفی تغییر دهید و نتایج آن را در گزارش بیاورید. مزایا و معایب این توابع فعالساز را نسبت به دیگری بررسی نمائید.
- نکته: از توابع فعالساز ReLU، TanH استفاده کنید.
- نکته: دقت نمائید در این آزمایش ها بقیه هایپرپارامترها ثابت هستند و از بهترین مدل قسمت (الف) استفاده نمائید.
- ج) تابع خطا شبکه را تغییر دهید و تاثیر تابع خطاهای متفاوت را در دقت آموزش شبکه بررسی نمائید، مجموعاً 2 مرتبه تابع خطا را تغییر دهید و نتایج آن را در گزارش بیاورید. دلیل این تفاوت را از منظر

ریاضی بررسی نمائید.

نکته :تابع خطا را از دو خانواده متفاوت انتخاب کنید ، به طور مثال تابع خطای Cross Entropy را با تابع خطای MSE می توانید مقایسه کنید.

نکته :دقت نمائید در این آزمایش ها بقیه هایپرپارامترها ثابت هستند و از بهترین مدل قسمت (ب) استفاده نمائید.

نکته :از بهینه سازهای Adam استفاده کنید. چرا در بعضی اوقات از SGD +momentum تنها استفاده می شود؟ معایب adam را شرح دهید.

نکته :دقت نمائید در این آزمایش ها بقیه هایپرپارامترها ثابت هستند و از بهترین مدل قسمت (ج) استفاده نمائید.

ه) با توجه به ارزیابی های انجام شده ، انتخاب کدام پارامترها بهترین نتیجه را می دهد؟
برای این مدل علاوه بر خطا و دقت ، معیارهای ارزیابی دیگر شامل Precision, Recall و F-Score را نیز گزارش نمائید.

و) اگر در یک مجموعه داده ، تعداد داده های دسته های ما با هم برابر نباشند چه مشکلی هنگام آموزش شبکه رخ می دهد؟ یک راه حل برای حل این مشکل ارائه دهید.

ز) از دسته های Airplane و Bird هر کدام نصف تصاویر داده آموزشی را انتخاب کنید و از بقیه دسته ها تمامی تصاویر داده آموزشی را انتخاب کنید و با استفاده از راه حل گفته شده در قسمت (و) مشکل نابرابری دسته ها را حل کنید و شبکه را آموزش دهید.

استفاده از CNN :

هدف در این تمرین افزایش دقت شبکه طبقه بند تصویر با اضافه کردن لایه های کانولوشنی و .. به شبکه است . در این سوال از کد قسمت قبل استفاده کنید و آن را توسعه دهید.

الف) به بهترین شبکه بدست آمده از قسمت قبل ، لایه های کانولوشنی را اضافه نمائید و شبکه را پیاده سازی نمائید. نتیجه بدست آمده را از نظر دقت و خطا با معماری MLP مقایسه نمائید.

ب) لایه های Batch normalization و Pooling را توضیح دهید و سپس این لایه ها را به توپولوژی شبکه اضافه نمائید و شبکه را پیاده سازی نمائید. نتیجه بدست آمده را از نظر دقت و خطا با معماری قسمت (الف) مقایسه نمائید.

ج) به معماری شبکه بدست آمده در قسمت (ب) اکنون Dropout را نیز اضافه نمائید و تاثیر آن را بررسی نمائید. چرا از Dropout در معماری شبکه عصبی استفاده می نمائیم؟

از جمله مشکلاتی که در بسیاری از موارد وجود دارد، محدود بودن دیتاست و یا نامتوازن بودن تعداد داده ها در هر کلاس است. این مشکلات سبب ایجاد اختلال در عملکرد شبکه می شود. روش های گوناگونی برای حل این مشکلات وجود دارد. یکی از این روش ها Data Augmentation است که در این سؤال با آن آشنا می شود.

1. در مورد Data Augmentation تحقیق کنید و توضیح دهید که نحوه عملکرد این روش چگونه است و تبدیل هایی که در آن استفاده می شود را شرح دهید. آیا از این روش برای داده های تست استفاده می شود؟ علت را شرح دهید.

2. با استفاده از چند تبدیل، از تصویر test.jpg، ۱۰ نمونه مصنوعی ایجاد کنید و آن ها را به همراه تصویر اصلی نمایش دهید.

3. همانند سؤال 2 از این تمرین، دیتاست Cifar10 را بارگذاری کنید سپس از کلاس های گربه و سگ در داده های آموزش، 4500 تصویر را حذف کنید. با این کار 90 درصد از تصاویر این دو کلاس حذف می شوند و تعداد تصاویر آموزش به 41000 می رسد. پس از حذف تصاویر، داده های جدید را با بهترین شبکه بدست آمده در سؤال 1 آموزش دهید. پس از آموزش مدل، ماتریس آشفتگی آن را برای داده های تست نمایش دهید. چه اختلالی در عملکرد شبکه رخ داده است؟

4. برای حل مشکل پیش آمده، با استفاده از Data Augmentation و ترکیبی از تبدیل های مختلف آن، تعداد تصاویر در کلاس های سگ و گربه را به 5000 برسانید تا در مجموع 50000 تصویر آموزش وجود داشته باشد. سپس داده های جدید را با بهترین شبکه به دست آمده در سؤال 1 آموزش و پس از آن ماتریس آشفتگی داده های تست را برای مدل آموزش دیده نمایش دهید. چه بهبودی نسبت به ماتریس آشفتگی بخش 3 ایجاد شده است؟ تفاوت ها با بخش 3 را شرح دهید.

نکته: برای Data Augmentation در این سؤال تنها مجاز به استفاده از ImageDataGenerator از کتابخانه Keras هستید.

-با پیشرفت های اخیر در استفاده از **CNN** در حوزه بینایی کامپیوتر ، مدل های معروفی از شبکه های عصبی کانولوشنی بوجود آمدند . در این سوال تصمیم گرفته شده است مدل های مختلفی را مطالعه و پیاده سازی نمایید. این تقسیم بندی بر اساس رقم آخر شماره دانشجویی شما انجام شده است.

نکته: با توجه به محدود بودن نسخه مدل ها در گزینه 8 و 9 دو مدل در این قسمت ها آمده است. در صورتی که یکی از گزینه های 8 یا 9 برای شما انتخاب شد ، لازم است در قسمت الف هر دو مدل را توضیح دهید و در قسمت (ب) تا (ه) فقط یک مدل را انتخاب و خواسته های این قسمت ها را انجام دهید.

model	number
EfficientNet	0
MobileNet	1
Inception	2
ResNet52	3
DenseNet	4
Xception	5
NASNet	6
VGG16	7
VGG19	8
SqueezeNet	9

الف (مختصرا موارد زیر را در مورد معماری انتخاب شده برای شبکه توضیح دهید:

- معماری شبکه
- توضیح نسخه های مختلف معماری و تفاوت آنها (در صورت وجود نسخه های مختلف)

- مزایا و معایب نسبت به سایر مدل ها
- محدودیت های ساینز تصویر ورودی
- پیش پردازش های اولیه برای تصویر ورودی
- ساینز خروجی مدل و معنای آن

(ب) شبکه انتخابی را با کمک **Transfer learning** پیاده سازی نمائید و سپس قسمت (ج) را اجرا نمائید.

(ج) ابتدا یک عکس رنگی با کیفیت مناسب از محیط اطراف خود بگیرید. (این عکس بایستی از دسته های قابل تشخیص توسط مدل باشد!) این عکس را در صورت نیاز پیش پردازش کنید، سپس به شبکه ای که پیاده سازی کرده اید داده و 3 دسته به ترتیب با بیشترین احتمال پیش بینی شده را در خروجی نشان دهید.

اضافه کردن **feature map** های لایه آخر و لایه اولیه برای رسم.

(د) اگر تصویر ورودی به شبکه جز اشیا قابل تشخیص توسط مدل نباشد، خروجی مدل قابل استناد نیست، راه حل شما برای حل این مشکل چیست؟ آن را پیاده سازی نمائید و نتیجه آن را گزارش نمائید.

(ه) یک لیست از اشیا قابل تشخیص توسط مدل بدست بیاورید، یک مجموعه داده جمع آوری کنید که شامل دو کلاس باشد (حداقل هر کلاس شامل 200 تصویر باشد، قاعدتا می توانید مجموعه داده بیشتری جمع آوری کنید و محدودیتی وجود ندارد) و مدل را با استفاده از این مجموعه داده مجددا آموزش دهید.

نکته: لازم است بنا به مدل انتخاب شده پیش پردازش های لازم روی مجموعه داده را انجام دهید و سپس با استفاده از وزن های مدل که قبلا آموزش داده است، مجددا شبکه را آموزش دهید.

نکته: اینکه چه دسته هایی جمع آوری می کنید و موضوع این دو دسته چیست کاملا بر عهده شماست اما تصاویر انتخابی باید مناسب طبقه بندی باشد. همچنین می توانید از دو دسته از دیتاست های معروف این حوزه همچون **ImageNet** نیز استفاده کنید.

نکته: دیتاست را در سامانه درس آپلود نکنید (فقط برای استفاده خودتان است). ولی توضیحاتی در ارتباط با دو دسته جمع آوری شده در گزارش بیاورید.

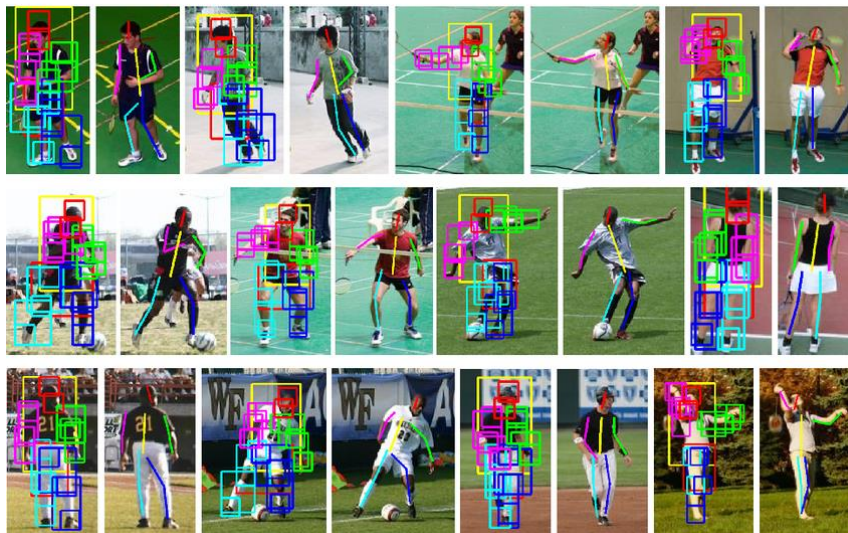
موارد زیر را در گزارش برای قسمت (ه) بیاورید:

در دو نمودار جداگانه تغییرات دقت و خطای مدل در هر دور را برای داده ی ارزیابی و آموزش نشان دهید.

Pose Estimation:

در این سوال میخواهیم با دادن تصویر به یک شبکه کانولوشنی محل سر، تنه و همچنین مفاصل دست ها و پاها را تعیین کنیم. بدین منظور از نسخه ساده دیتاست LSP استفاده می کنیم که از ۲۰۰۰ تصویر تشکیل شده است و محل نقاط هدف به صورت یک فایل ماتریس متلب در آن داده شده است این دیتاست در گروه تلگرام و پوشه تمرین قرار داده شده است. توجه داشته باشید که فایل ماتریس متلب joints.mat را میتوان به کمک کتابخانه های موجود در پایتون خواند. همچنین در این تمرین نیازی به استفاده از پوشه visualized نیست.

70 درصد داده ها را برای آموزش، 15 درصد برای ولیدیشن و 15 درصد را برای تست بکار بگیرید. در شکل زیر نمونه ای از تصاویر موجود در این دیتاست مشاهده میشود



شبکه ی معرفی شده در مقاله ی زیر برای این سوال قابل استفاده است:

<https://arxiv.org/abs/1312.4659>

شبکه معرفی شده در مقاله از دو قسمت **Initial stage** و **Stages** تشکیل شده است. با این هدف که در شبکه ی اول یک تخمین اولیه از مکان مفصل ها زده میشود و هدف شبکه های بعدی دقیقتر کردن این تخمین است.

در این قسمت شما باید علاوه بر خواندن مقاله و آماده سازی داده ها و برچسب های آنها شبکه **initial stage** را پیاده سازی کنید و منحنی **LOSS** آموزش، ولیدیشن و تست و همچنین مقادیر معیار های **PCP** و **PDJ** را برای هر ۱۴ مفصل رسم کنید . همچنین به عنوان نمونه مکان مفصل را مانند شکل برای ۴ تصویر از داده های تست رسم کنید.