دانشگاه صنعتی خواجه نصیرالدین طوسی
دانشکدهٔ مهندسی برق

# درس رباتیک و بینایی ماشین

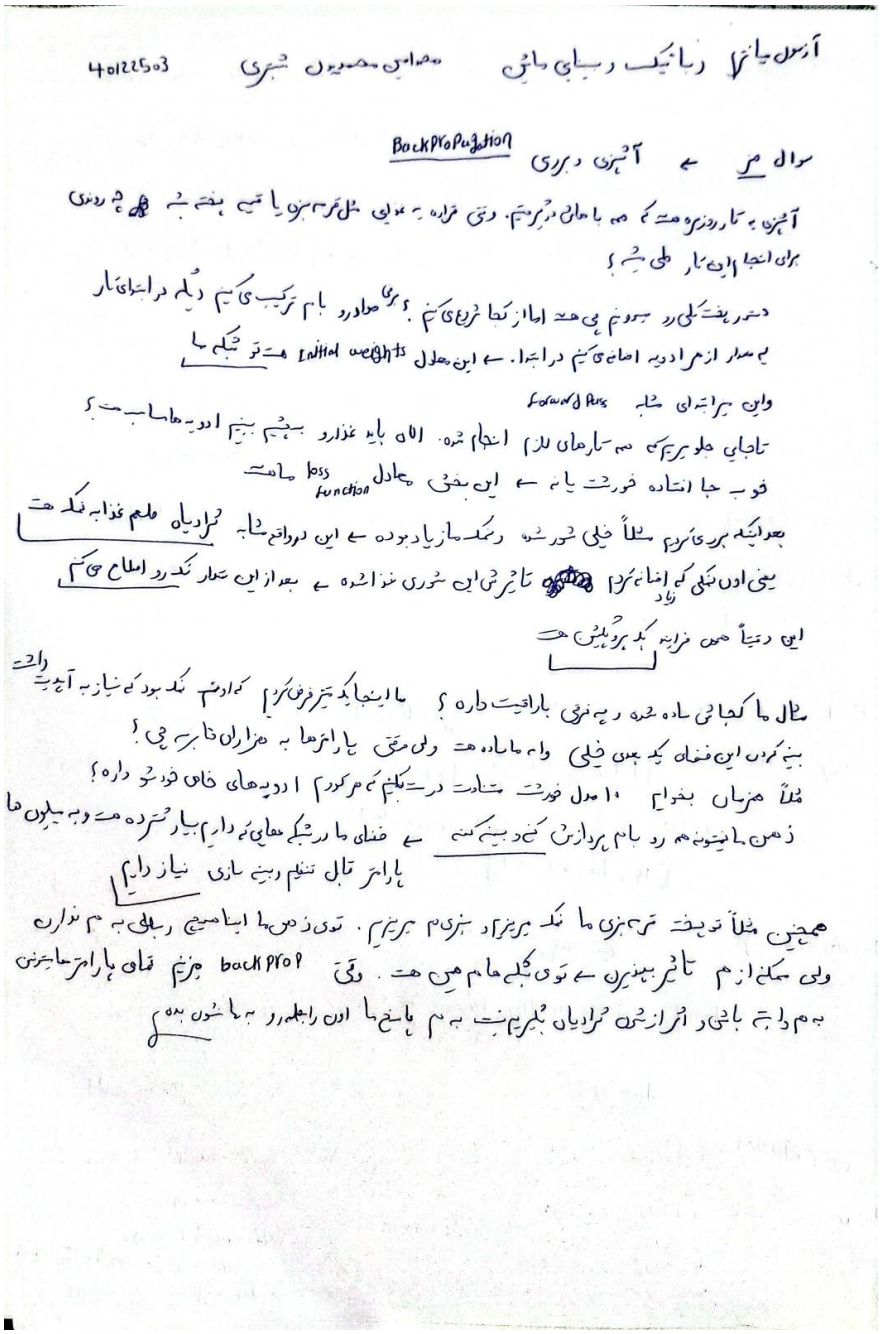
استاد: دکتر حمیدرضا تقی راد

استاد: دکتر محمدجواد احمدی

آزمون میانترم

| | |
|---|---|
| نام و نام خانوادگی | محمدامین محمدیون شبستری |
| شمارهٔ دانشجویی | ۴۰۱۲۲۵۰۳ |
| تاریخ | خرداد ۱۴۰۵ |

## ۱ سوال ۰



آزمون میانترم رباتیک و بینایی ماشین محمدامین محمدیون شبستری 40122503

سوال ۳ ← آشنایی و بررسی BackPropagation

آشنایی به تا روزی هست که همه با مدل درگیر هستیم. وقتی قراره به مدلی مثل ترسیم یا تعیین بخش چه روندی برای انجام این کار طی بشه؟

دستور پخت کلی رو میدونم چی هست اما از کجا شروع کنم؟ برای مواردی با هم ترکیب می‌کنیم ولی در ابتدای کار یه مقدار از مراد دیه اضافه می‌کنیم در ابتدا. ← این مقدار initial weights هست تو شکلها

و این پیرایه‌ای مشابه Forward Pass تا جایی جلو بریم که همه سازهای لازم انجام شده. اما باید غذا رو بچشیم ببینیم ادویه‌ها مناسب هست یا خوب جا افتاده خورشت یا نه ← این بخش معادل loss function ماهست

بعد اینکه بررسی کردیم مثلاً خیلی شور شده و نمک ما زیاد بوده ← این در واقع مشابه گرادیان فلعم غذا به نمک هست یعنی اون نمکی که اضافه کردم زیاد تاثیری این شوری نداشته ← بعد از این مقدار نمک رو اصلاح می‌کنم

این دقیقاً همون فرایند گرادیان هست

مثال ما کجایی ساده شده و به نوعی با واقعیت داره؟ ما اینجا یک تیر فرض کردیم که اومدیم نمک بود که نیاز به آبپز داشت بنی کردن این فضا یک بعدی خیلی واقع‌بینانه ماساده هست ولی وقتی پارامترها به هزاران تا برسه چی؟ مثلاً همزمان بخوام ۱۰ مدل خورشت مشابه درست بکنم که هر کدوم ادویه‌های خاص خودشو داره؟ ذهن ما نمیتونه همه رو با هم پردازش کنه و ببینه کجا ← فضای ما در شکل‌هایی که داریم بسیار گسترده هست و به میلیون‌ها پارامتر قابل تنظیم و بهینه‌سازی نیاز داریم

همچنین مثلاً تو پخت قرمه‌سبزی ما نمک بریزیم و سبزی هم بریزیم. توی ذهن ما اینا هیچ ربطی به هم ندارن ولی ممکنه از هم تاثیر بپذیرن ← توی شکلها هم همین هست. وقتی backprop چیزی تمام پارامترها رو به هم وابسته باشن و اگر از شون گرادیان بگیریم به هم پاسخ‌ها اون رابطه رو به ما نشون بده

شکل ۱

# سوال ۱

## الف) رابطه کلی ماتریس تبدیل دوربین به پایه ($^0T_C$)

با توجه به ویژگی ضرب زنجیره‌ای ماتریس‌های تبدیل همگن، رابطه کلی به صورت زیر است:

$$^0T_C = ^0T_E \cdot ^ET_C \tag{۱}$$

ماتریس $^ET_C$ بر اساس دوران $90°$ حول محور $X$ و انتقال $d = [0, 0, 5]^T$ به شکل زیر تشکیل می‌شود:

$$^ET_C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(90°) & -\sin(90°) & 0 \\ 0 & \sin(90°) & \cos(90°) & 5 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{۲}$$

رابطه نهایی کالیبراسیون:

$$^0T_C = ^0T_E \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{۳}$$

## ب) مراحل محاسباتی و فرمول ساختار ضرب برای یافتن $^0P$

ابتدا بردار مختصات شیء در قاب دوربین ($^CP = [1, 10, -1, 2]^T$) را با افزودن درایه ۱ به فرم همگن در می‌آوریم:

$$^CP_{homogeneous} = \begin{bmatrix} 1 \\ 10 \\ -1 \\ 2 \\ 1 \end{bmatrix} \tag{۴}$$

فرمول ساختار زنجیره‌ای ضرب ماتریس‌ها برای انتقال مختصات به قاب پایه ($^0P$) به صورت زیر است:

$$^0P_{homogeneous} = ^0T_C \cdot ^CP_{homogeneous} = ^0T_E \cdot ^ET_C \cdot ^CP_{homogeneous} \tag{۵}$$

با جایگذاری مقادیر عددی، ساختار دقیق ضرب عبارت است از:

$$^0P_{homogeneous} = ^0T_E \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 10 \\ -1 \\ 2 \\ 1 \end{bmatrix} \tag{۶}$$

# سوال ۲

## ۱) تأثیر اشتباه بر ماتریس‌های انتقال $A_i$

در قرارداد Standard ،DH ترم‌های متغیر مفصلی ($\theta_i$) و انتقال محوری ($d_i$) بر اساس محور $Z$ تعریف می‌شوند. قرار دادن محور $Z$ در راستای طول لینک به جای محور مفصل، پیامدهای زیر را دارد:

- جابجایی هندسی پارامترها: پارامتر $a_i$ (طول لینک) که باید به عنوان انتقال در راستای $X_i$ اعمال شود، به عنوان انتقال در راستای $Z$ ظاهر شده و جای ترم‌های مربوط به $a_i$ و $d_i$ درون ماتریس همگن $A_i$ عوض می‌شود.
- توصیف نادرست دوران: ماتریس ریاضی بر پایه چرخش حول $Z$ است. با تغییر جهت این محور، فرمول‌ها دیگر قادر به توصیف درست حرکت دورانی واقعی مفصل ربات نیستند.

## ۲) دلیل تأکید DH بر انطباق محور $Z$ با محور مفصل (Joint Axis)

- تطابق با متغیرهای مفصلی: برای اینکه رفتار ریاضی مدل با متغیرهای فیزیکی موتورها (زوایای $\theta_i$ برای مفاصل دورانی) کاملاً همگام باشد، قرارداد DH محرکِ مستقلِ هر درجه آزادی را منحصراً روی محور $Z$ آن مفصل تعریف می‌کند.
- جداسازی روابط حرکتی :(Decoupling) این کار مانع از جفت‌شدن (Coupling) پیچیده درجات آزادی می‌شود. در صورت عدم رعایت، تغییر یک زاویه مفصل به جای یک دوران ساده، به صورت ترکیبی ناخواسته از دوران و انتقال‌های کاذب در ماتریس دیده خواهد شد.

## ۲ سوال ۳

سوال 3 ← GHE سعی میکند تا کنتراست کل تصویر را یکسان کند ← سایه هایش از حد روشن میشوند

نوای روشن هم خیلی زیاد روشن میشود

بعد نیز تاملی شدن اشباع میشود

در نتیجه ترکه های نیز کاملاً محو میشوند و قابل تشخیص نخواهند بود

روش CLAHE ← تصویر را به بلوک های مختلف تقسیم میکند عملیات را روی هر بلوک مستقل از بلوک دیگر انجام میدهد

با این کار اگر یک بلوک داخل سایه باشد تقلیل هیستوگرام براساس همین بلوک و محیط سایه ای که وجود دارد انجام میشود

در نهایت هم به همه بلوک ها با درونیابی خطی پیوند می خورند تا حالت یکنواخی رعایت شدن پیش نیاید

پارامتر clip limit ← در هر بلوک یک سقف و محدودیت برای ارتفاع هیستوگرام تعیین وجود. اگر تعداد پیکسل های یک شدت روشنایی از این سقف عبور کند مقدار اضافی clip میشود به طور یکنواخت بین سطوح خاکستری این بلوک پخش میشود

پارامتر Grid size ← ابعاد هر کدام از بلوک ها را مشخص میکند ← بلوک کوچک باشد ← کنتراست محلی بالا میرود و نویزهای فرکانسی بالا مثل نویزهای سطح فلز تقویت میشود

اگر بلوک بزرگ باشد ← به سمت الگوریتم GHE میرود از نظر رفتاری

پس انتخاب بین این پنجره ابعاد باید به شکل تقویت نویز بخورد و نه به سمت GHE حرکت کند

شکل ۲

# سوال ۴

## ۱) تعریف ریاضی فیلتر Sobel در راستای افقی ($G_x$)

فیلتر لبه‌یاب Sobel در راستای افقی (که وظیفه سنجش گرادیان و تغییرات شدت روشنایی در امتداد محور افقی یا همان ستون‌ها را دارد)، به صورت یک ماتریس $3 \times 3$ استاندارد زیر تعریف می‌شود:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{۷}$$

## ۲) محاسبات دستی کانولوشن برای پیکسل مرکزی

ماتریس تصویر ورودی $I$ با ابعاد $3 \times 3$ به صورت زیر داده شده است:

$$I = \begin{bmatrix} 10 & 10 & 100 \\ 10 & \mathbf{10} & 100 \\ 10 & 10 & 100 \end{bmatrix} \tag{۸}$$

هدف، محاسبه مقدار خروجی فیلتر برای پیکسل مرکزی (در موقعیت سطر دوم و ستون دوم با مقدار 10) بدون در نظر گرفتن Padding است. عملگر کانولوشن (ضرب نقطه‌ای درایوهای متناظر ماسک فیلتر بر روی همسایگی تصویر و مجموع آن‌ها) به صورت زیر محاسبه می‌شود:

$$\text{Output} = \sum_{i=1}^{3} \sum_{j=1}^{3} (I_{ij} \times G_{x,ij})$$

با جایگذاری مقادیر ماتریس تصویر $I$ و فیلتر $G_x$، محاسبات سطر به سطر به شرح زیر است:

$$\begin{aligned} \text{Output} = \big(10 \times (-1) + 10 \times 0 + 100 \times 1\big) & \quad \leftarrow \text{سطر اول} \\ + \big(10 \times (-2) + 10 \times 0 + 100 \times 2\big) & \quad \leftarrow \text{سطر دوم} \\ + \big(10 \times (-1) + 10 \times 0 + 100 \times 1\big) & \quad \leftarrow \text{سطر سوم} \end{aligned}$$

حاصل عددی هر سطر عبارت است از:

- سطر اول: $-10 + 0 + 100 = 90$
- سطر دوم: $-20 + 0 + 200 = 180$
- سطر سوم: $-10 + 0 + 100 = 90$

مجموع نهایی برای پاسخ پیکسل مرکزی:

$$\text{Output} = 90 + 180 + 90 = 360 \tag{۹}$$

بنابراین، مقدار خروجی فیلتر برای پیکسل مرکزی برابر 360 است.

## ۳) تفسیر مفهومی نتیجه در خصوص جهت و شدت گرادیان لبه

خروجی حاصل شده (360) حاوی اطلاعات فیزیکی و هندسی زیر از ساختار تصویر در این همسایگی است:

۱. شدت گرادیان لبه (Edge Magnitude): مقدار عددی بزرگ به‌دست‌آمده (360) بیانگر وجود یک لبه‌ی بسیار تیز و با تباین (کنتراست) بالا در این ناحیه است که ناشی از جهش ناگهانی شدت روشنایی از غلظت 10 به 100 در ستون سوم تصویر می‌باشد.

۲. جهت لبه و گرادیان (Edge Gradient & Direction):

- امتداد هندسی لبه: با توجه به اینکه مقادیر ستون‌های اول و دوم همگی 10 و ستون سوم همگی 100 هستند، تغییرات موازی سطرها رخ داده و خود خطِ لبه به صورت یک لبه‌ی کاملاً عمودی (Vertical Edge) ظاهر شده است.
- جهت بردار گرادیان: علامت مثبت خروجی (+360) نشان می‌دهد که بردار گرادیان (جهت بیشترین افزایش شدت روشنایی) دقیقاً در امتداد محور افقی مثبت و از چپ به راست (زاویه $0°$) است؛ یعنی تصویر با حرکت به سمت راست، به شدت روشن‌تر می‌شود.

# سوال ۵

## مقدمه و تعریف ریاضی تابع فعال‌ساز سیگموید

تابع فعال‌ساز سیگموید (Sigmoid) به صورت زیر تعریف می‌شود:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{۱۰}$$

مشتق این تابع نسبت به ورودی آن، رابطه بسیار معروفی دارد که برحسب خود تابع خروجی بیان می‌شود:

$$\sigma'(x) = \sigma(x)\big(1-\sigma(x)\big) \tag{۱۱}$$

$$\frac{df}{ds} = 1 - 2s = 0 \implies s = 0.5$$

با جایگذاری $s = 0.5$ در رابطه مشتق، حداکثر مقدار آن به دست می‌آید:

$$\max\big(\sigma'(x)\big) = 0.5 \times (1-0.5) = 0.25 \tag{۱۲}$$

بنابراین، برای هر ورودی دلخواه $x$، همواره رابطه زیر برقرار است:

$$0 < \sigma'(x) \leq 0.25 \tag{۱۳}$$

## اثبات مفهومی پدیده گرادیان محو شونده (Vanishing Gradient) با قاعده زنجیره‌ای

$$\frac{\partial J}{\partial w^{(1)}} = \frac{\partial J}{\partial a^{(L)}} \cdot \frac{\partial a^{(L)}}{\partial z^{(L)}} \cdot \frac{\partial z^{(L)}}{\partial a^{(L-1)}} \cdots \frac{\partial a^{(2)}}{\partial z^{(2)}} \cdot \frac{\partial z^{(2)}}{\partial a^{(1)}} \cdot \frac{\partial a^{(1)}}{\partial z^{(1)}} \cdot \frac{\partial z^{(1)}}{\partial w^{(1)}} \tag{۱۴}$$

از آنجا که $\frac{\partial a^{(l)}}{\partial z^{(l)}} = \sigma'(z^{(l)})$ و $\frac{\partial z^{(l)}}{\partial a^{(l-1)}} = w^{(l)}$ است، رابطه فوق را می‌توان به شکل زیر خلاصه کرد:

$$\frac{\partial J}{\partial w^{(1)}} = \frac{\partial J}{\partial a^{(L)}} \cdot \left(\prod_{i=2}^{L} w^{(i)} \cdot \sigma'(z^{(i)})\right) \cdot \sigma'(z^{(1)}) \cdot a^{(0)} \tag{۱۵}$$

تحلیل صفر شدن سیگنال خطا در لایه‌های اول:

۱. اثر حاصل‌ضرب متوالی اعداد کوچک: از آنجا که مقدار $\sigma'(z)$ همواره کوچکتر یا مساوی $0.25$ است، در یک شبکه عمیق با افزایش تعداد لایه‌ها، عبارت $\prod \sigma'(z^{(i)})$ حاصل‌ضربی از تعداد زیادی عدد کوچک (کمتر از $0.25$) خواهد بود. به عنوان مثال، تنها برای ۵ لایه متوالی، این اثر مرتبه‌ای در حدود $(0.25)^5 \approx 0.00097$ ایجاد می‌کند.

۲. مکانیزم اشباع :(Saturation) اگر ورودی لایه‌ها ($z$) مقادیر خیلی بزرگ یا خیلی کوچک به خود بگیرند، مقدار مشتق سیگموید به شدت به سمت صفر میل می‌کند ($\sigma'(x) \to 0$).

در نتیجه این دو عامل، با حرکت سیگنال خطا از لایه‌های انتهایی به سمت لایه‌های ابتدایی، مقدار گرادیان به صورت نمایی کاهش یافته و در لایه‌های اول عملاً **صفر** می‌شود. وقتی گرادیان صفر شود، وزن‌های لایه‌های ابتدایی به‌روزرسانی نشده و فرآیند آموزش شبکه‌های عمیق متوقف می‌گردد.

## راهکارهای مدرن در معماری شبکه‌های CNN برای رفع این مشکل

معماری‌های مدرن بینایی ماشین (CNN) از چندین تکنیک کلیدی برای حفظ فرآیند جریان گرادیان در لایه‌های عمیق استفاده می‌کنند:

۱. استفاده از تابع فعال‌ساز ReLU و مشتقات آن: به جای توابع محدودکننده مثل Sigmoid، از تابع $f(x) = \max(0, x)$ استفاده می‌شود. مشتق این تابع برای ورودی‌های مثبت همواره دقیقاً برابر 1 است ($\sigma'(x) = 1$). این ویژگی باعث می‌شود که سیگنال گرادیان بدون تضعیف و کاهش نمایی از لایه‌ها عبور کند.

۲. اتصالات باقیمانده (Residual Connections): Skip / Connections در معماری‌هایی مانند ResNet، از اتصالات میان‌بر استفاده می‌شود که ورودی یک لایه را مستقیماً به چند لایه جلوتر جمع می‌کند ($y = F(x) + x$). در هنگام پس‌انتشار، مشتق این رابطه ترم $+1$ را به جریان گرادیان اضافه می‌کند:

$$\frac{\partial y}{\partial x} = \frac{\partial F(x)}{\partial x} + 1$$

این مقدار 1 تضمین می‌کند که حتی اگر گرادیان مسیر اصلی ($F(x)$) صفر شود، سیگنال خطا از طریق مسیر میان‌بر بدون هیچ مانعی مستقیماً به لایه‌های ابتدایی منتقل شود.

۳. لایه نرمال‌سازی دسته‌ای (Batch Normalization): این لایه ورودی هر لایه را به گونه‌ای نرمال‌سازی می‌کند که میانگین صفر و واریانس یک داشته باشند. این کار مانع از بزرگ شدن بیش از حد یا کوچک شدن ورودی‌ها شده و از ورود توابع فعال‌ساز به نواحی اشباع (جایی که مشتق صفر است) جلوگیری می‌کند.

## سوال ۶

### الف) محاسبه ابعاد فضایی (ارتفاع و عرض) تانسور خروجی

برای محاسبه ابعاد فضایی خروجی یک لایه کانولوشن دوبعدی ($H_{out}$ و $W_{out}$)، از فرمول استاندارد زیر استفاده می‌کنیم:

$$N_{out} = \left\lfloor \frac{N_{in} - K + 2P}{S} \right\rfloor + 1 \tag{۱۶}$$

که در آن پارامترهای داده شده در مسئله عبارتند از:

- ابعاد فضایی ورودی ($N_{in}$): 128 (با توجه به تانسور ورودی $3 \times 128 \times 128$)
- ابعاد فیلتر ($K$ یا Kernel Size): 5
- حاشیه ($P$ یا Padding): 2
- گام حرکت ($S$ یا Stride): 2

با جایگذاری این مقادیر در فرمول، محاسبات به صورت زیر انجام می‌شود:

$$\begin{aligned} H_{out} = W_{out} &= \left\lfloor \frac{128 - 5 + 2(2)}{2} \right\rfloor + 1 \\ &= \left\lfloor \frac{128 - 5 + 4}{2} \right\rfloor + 1 \\ &= \left\lfloor \frac{127}{2} \right\rfloor + 1 \\ &= \lfloor 63.5 \rfloor + 1 \\ &= 63 + 1 = 64 \end{aligned}$$

نتیجه بخش الف: ابعاد فضایی تانسور خروجی برابر با $64 \times 64$ است (ابعاد کامل تانسور خروجی با احتساب تعداد کانال‌ها $64 \times 64 \times 64$ خواهد بود).

### ب) تعداد کل پارامترهای قابل آموزش (Trainable Parameters) با احتساب Bias

فرمول کلی برای محاسبه تعداد پارامترهای قابل آموزش در یک لایه کانولوشن استاندارد عبارت است از:

$$\text{Parameters Total} = (K_H \times K_W \times C_{in} \times C_{out}) + C_{out} \tag{۱۷}$$

که در آن پارامترها به شرح زیر تعریف می‌شوند:

- $K_H \times K_W$ (ابعاد فیلتر): $5 \times 5$
- $C_{in}$ (تعداد کانال‌های ورودی): 3
- $C_{out}$ (تعداد کانال‌های خروجی / تعداد فیلترها): 64

- $C_{out}$ (تعداد لایه‌های بایاس): 64 (به تعداد کانال‌های خروجی، بیایس آموزش‌پذیر داریم)

مراحل محاسبه عددی:

۱. تعداد وزن‌ها :(Weights)

$$5 \times 5 \times 3 \times 64 = 25 \times 3 \times 64 = 75 \times 64 = 4800$$

۲. تعداد بایاس‌ها :(Biases) 64

۳. مجموع کل پارامترها:

$$\text{Parameters Total} = 4800 + 64 = 4864$$

نتیجه بخش ب: تعداد کل پارامترهای قابل آموزش در این لایه برابر با 4864 است.

## ۳ سوال ۷

سوال 7 <= خروجی کانولوشن

stride default=1

$P=1 \quad K=3 \quad out_c=16 \quad in_c=3$

self.conv1

$$w_{out}=\left[\frac{in_c-K+2P}{S}\right]+1 \Rightarrow \left[\frac{64-3+2}{1}\right]+1=64$$

ابعاد تنسور بعد از کانولوشن و تابع فعال‌ساز Relu

[B, 16, 64, 64]

خروجی self.pool

$$w_{pool}=\frac{w_{conv}}{2}=32$$

[B, 16, 32, 32]

محاسبه ابعاد flatten شده ←

channel × width × height = 16×32×32 = 16384

الف) مقدار صحیح برای view برابر است با 16×32×32

ب) برای اینکه حساس نباشد به ابعاد بهترین کار استفاده از adaptive AvgPool2d قبل از flatten کردن

باید توی init تعریف کنیم

self.adaptive_pool = nn.adaptiveAvgPool2d(16,16)

کد اصلاح شده

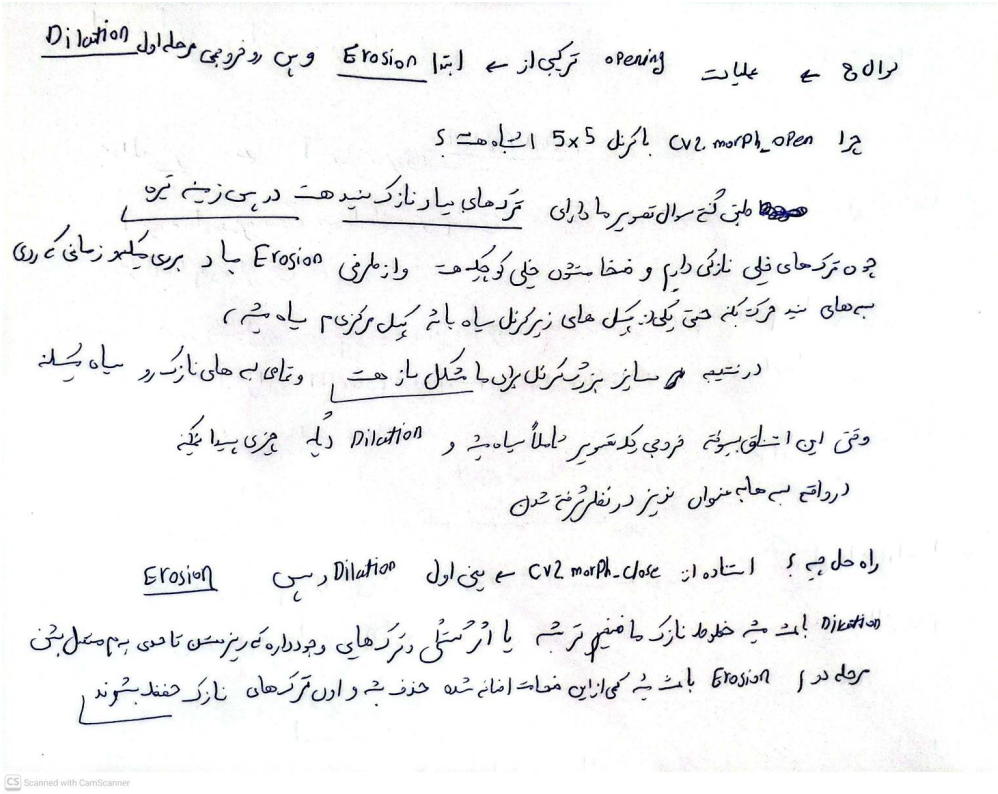
```
def forward (self,x):
    x = self.pool (F.relu ( self.conv1(x))
    x = self.adaptive_pool (x)
    x= x.view (-1 , 16*16*16)
    x=self.fc1 (x)
```

الان هر چی باشد ابعاد ورودی لایه fc به 16×16 تبدیل می‌شود

Scanned with CamScanner

شکل ۳

## ۴ سوال ۸



برای ۵ ← عملیات opening ترکیبی از ← ابتدا Erosion و پس روند فرو بعد مرحله اول Dilation

چرا cv2.morph_open با کرنل 5x5 ایجاد است؟

طبق گفته سوال تصویر ما دارای ترک‌های بسیار نازک سفید است در پس زمینه تیره

چون ترک‌های خیلی نازکی داریم و فضا ستون خیلی کوچک است و از طرفی Erosion باعث بزرگ‌تر شدن زمانی که روی

سفیدها سفید ترک کنه حتی یکی از پیکسل های زیر کرنل سیاه باشه پیکسل مرکزی هم سیاه میشه

در نتیجه سایز بزرگ کرنل برای ما مشکل ساز است و تمامی سفیدهای نازک رو سیاه میکنه

وقتی این اتفاق بیوفته فرقی یک تصویر کاملاً سیاه میشه و Dilation دیگه چیزی پیدا نمیکنه

در واقع سفیدها عنوان نویز در نظر گرفته شدن

راه حل چیه؟ استفاده از cv2.morph_close ← یعنی اول Dilation و سپس Erosion

باعث میشه خطوط نازک یا ضعیف تر بشه یا اگر گسستگی در ترک‌های وجود داره که ریز هستن تا حدی به هم متصل بشن

مرحله دوم Erosion باعث میشه کمی از این ضخامت اضافه شده حذف بشه و اون ترک‌های نازک حفظ بشوند

Scanned with CamScanner

شکل ۴