



دانشگاه صنعتی اصفهان
دانشکده مهندسی برق و کامپیوتر

عنوان: تکلیف سوم درس یادگیری عمیق

نام و نام خانوادگی: علیرضا ابره فروش

شماره دانشجویی: ۹۸۱۶۶۰۳

نیم سال تحصیلی: پاییز ۱۴۰۲

مدرس: دکتر سمانه حسینی سمنانی

دستیاران آموزشی: مریم محمدی-علی بزرگ زادارباب

استفاده از تصاویر در شبکه‌های عصبی معمولی (MLP) دارای مشکلاتی بود که با ظهور شبکه‌های عصبی کانولوشنی (CNN) حل شد:

۱. تعداد پارامترها و اتصالات:

در MLP، هر نورون لایه ورودی با تمام نورون‌های لایه خروجی متصل بود. این اتصالات منجر به افزایش شدید تعداد پارامترها می‌شد، که با تعداد بزرگ تصاویر (از نظر تعداد پیکسل) به سرعت غیرقابل مدیریت می‌شد. در CNN، از لایه‌های کانولوشنی برای به اشتراک‌گذاری ویژگی‌ها در تصاویر استفاده می‌شود. این لایه‌ها با اعمال فیلترها (کرنل‌ها) به تصویر، ویژگی‌های مختلف را استخراج می‌کنند و این باعث کاهش چشم‌گیر تعداد پارامترها و اتصالات شبکه می‌شود. به عبارت دیگر، CNN با استفاده از به اشتراک‌گذاری ویژگی‌ها، تعداد پارامترها را به شدت کاهش می‌دهد.

۲. مقاومت به تغییرات مکانی:

MLP به طور کامل حساس به تغییرات مکانی در تصویر بود. به عبارت دیگر، اگر یک الگو یا ویژگی در یک مکان خاص در تصویر وجود داشت، MLP قادر به تشخیص آن در سایر نقاط تصویر نبود. CNN با استفاده از لایه‌های کانولوشنی و استفاده از فیلترها، از این قابلیت برای شناسایی ویژگی‌ها در تمام تصویر به خوبی استفاده می‌کند. به این ترتیب، شبکه CNN مقاوم‌تر به تغییرات مکانی در تصاویر می‌شود و توانایی خود را در تشخیص الگوها و ویژگی‌ها در مکان‌های مختلف تصویر افزایش می‌دهد.

۳. عدم حفظ ساختار مکانی: MLP، ساختار مکانی تصویر حذف می‌شد و هر پیکسل به عنوان یک ویژگی مستقل در نظر گرفته می‌شد. این عدم حفظ ساختار مکانی اطلاعات مهمی را از دست می‌داد، به خصوص برای تصاویری که الگوها و اطلاعات مکانی مهمی دارند. CNN با استفاده از لایه‌های کانولوشنی و اعمال فیلترها، قابلیت حفظ ساختار مکانی را به شبکه اضافه کرده و اجازه می‌دهد تا ویژگی‌ها به صورت محلی در تصویر استخراج شوند. این امکان باعث می‌شود که شبکه بتواند اطلاعات مکانی را مورد توجه قرار داده و الگوهای مکانی پیچیده‌تری را در تصاویر تشخیص دهد.

۴. تفسیرپذیری پایین:

MLPها به دلیل تعداد بسیار زیاد پارامترها، به صورت کلی دارای تفسیرپذیری پایین بودند. یعنی مشخص کردن دقیق اینکه شبکه چگونه تصمیمات خود را اتخاذ کرده است، معمولاً مشکل بود. CNN با تعداد کمتر پارامترها و استفاده از فیلترها، قابلیت تفسیرپذیری بیشتری دارد. این به این معناست که می‌توان بهتر فهمید که شبکه در تصمیم‌گیری‌های خود چگونه از ویژگی‌ها استفاده می‌کند.

۵. حساسیت به اندازه تصویر:

MLPها به صورت ثابت و بدون توجه به ابعاد تصویر (مثلاً ابعاد ورودی ثابت دارای تعداد ثابت نورون‌ها) عمل می‌کردند. این باعث می‌شد که اگر تصویر ورودی ابعاد متفاوتی داشته باشد، شبکه به طور مستقیم با آن کار نکند. CNN با استفاده از لایه‌های کانولوشنی می‌تواند به تصاویر با ابعاد مختلف و با استفاده از فیلترهای متفاوت به خوبی پاسخ دهد و حساسیت کمتری نسبت به ابعاد تصویر نشان دهد.

۶. برخورد با تعداد زیاد پارامترها:

با افزایش اندازه تصاویر، تعداد پارامترهای مورد نیاز برای یک MLP به صورت نمایی افزایش می‌یابد. این موضوع باعث ایجاد مدل‌های بسیار پیچیده و سنگین می‌شود که دشواری در آموزش، نگهداری و استفاده از آن‌ها را افزایش می‌دهد.

۷. عدم توانایی در مدل سازی ویژگی های سلسله مراتبی:

MLP به طور مستقیم قادر به مدل سازی ویژگی های سلسله مراتبی و پیچیده تصاویر نیستند. به عبارت دیگر، آن ها قادر به استخراج ویژگی های موقعیت مکانی، الگوها و ساختارهای سلسله مراتبی نیستند که در بینایی ماشین بسیار مهم است.

۸. حساسیت به تغییرات شدت نور و ظروف نوری:

MLP در مقابل تغییرات شدت نور و ظروف نوری حساس هستند. این به این معناست که تصاویر با نور متفاوت یا ظروف نوری متفاوت ممکن است تأثیر زیادی بر عملکرد MLP داشته باشند.

۹. اشتباه زهای غیرقابل کنترل:

به دلیل تعداد بالای پارامترها و عدم استفاده از الگوهای مکانی، MLP ها ممکن است به آموزش بیش از حد به داده ها وابسته شوند و اشتباه زهای غیرقابل کنترل در عملکرد آن ها ایجاد شود.

۱۰. کارایی ضعیف در مسائل تشخیص الگو:

در مسائل تشخیص الگو و ویژگی های پیچیده در تصاویر، MLP ها عملکرد ضعیفی دارند. زیرا این مدل ها نمی توانند ویژگی های سلسله مراتبی و پیچیده را به صورت کامل مدل کنند.

شبکه های عصبی کانولوشنی (CNN) با توجه به مزایایی که در پردازش تصاویر دارند، این مشکلات را به حداقل می رسانند و بهبودهای مهمی را در زمینه بینایی ماشین و پردازش تصویر به ارمغان آورده اند.

۲

۱. استفاده از لایه های کانولوشنی:

در بینایی کامپیوتر کلاسیک، اغلب از فیلترها و عملیات پردازش سیگنال سنتی برای استخراج ویژگی ها استفاده می شود. این روش ها معمولاً به صورت دستی تنظیم می شوند و برای ویژگی های خاص استفاده می شوند. در CNN، از لایه های کانولوشنی استفاده می شود که خودشان یک نوع فیلتر هستند. این لایه ها با اعمال فیلترها به تصویر، ویژگی های مختلف را به صورت خودکار و سلسله مراتبی استخراج می کنند. این عمل باعث افزایش توانایی شبکه در تشخیص ویژگی های مکانی و سلسله مراتبی می شود.

۲. پارامترها و اتصالات کمتر:

در بینایی کامپیوتر کلاسیک، تعداد زیادی از پارامترها به صورت دستی تنظیم می شوند و شبکه ها ممکن است به سرعت پیچیده شوند. در CNN، با به اشتراک گذاری ویژگی ها (استفاده از فیلترها)، تعداد پارامترها به شدت کاهش می یابد. این باعث می شود شبکه قابل مدیریت تر باشد و از اورفیتینگ (overfitting) کاسته شود.

۳. حفظ ساختار مکانی:

بینایی کامپیوتر کلاسیک عمدتاً از روش هایی استفاده می کند که ساختار مکانی تصویر را حذف می کند و به هر پیکسل به عنوان یک واحد مستقل نگاه می کند. CNN با لایه های کانولوشنی خود، قابلیت حفظ ساختار مکانی تصویر را فراهم می کند. این به این معناست که شبکه قادر است به ویژگی ها و الگوهای موقعیت مکانی توجه داشته باشد و این امکان را دارد که اطلاعات مکانی در تصویر را بهتر به کار بگیرد.

۴. آموزش انتقالی:

CNN از آموزش انتقالی بهره می‌برد که به انتقال وزن‌های یک مدل آموزش دیده برای حل یک مسئله خاص می‌پردازد. این به شبکه‌ها این امکان را می‌دهد که از دانش کسب‌شده در یک مسئله به مسائل مشابه دیگر نیز استفاده کنند.

۵. نیاز به استخراج دستی ویژگی‌ها تصویر توسط یک متخصص:

یکی از مشکلات عمده الگوریتم‌های یادگیری ماشین اولیه که در اوایل دهه ۱۹۶۰ به وجود آمد، نیاز به استخراج دستی و دقیق ویژگی‌ها از تصاویر توسط یک متخصص بود. حتی اگر اندکی اتوماسیون نیز در آن زمان مورد استفاده قرار می‌گرفت، نیاز به تنظیم دقیق توسط یک متخصص وجود داشت. زیرا الگوریتم‌هایی مانند SVM یا KNN برای یافتن ویژگی‌های مهم به کار می‌رفتند. این نیاز به ساخت یک مجموعه داده با ویژگی‌های مشخص برای مدل برای یادگیری از آن می‌رفت. بنابراین، تکنیک‌های یادگیری عمیق یک کمک بزرگ برای افراد تخصصی بود. چرا که دیگر نیازی به نگرانی درباره انتخاب دستی ویژگی‌ها برای یادگیری مدل نداشتند.

۶. نیاز به منابع محاسباتی سنگین:

یادگیری عمیق یک وظیفه با وزن سنگین محاسباتی است که به همین دلیل در دهه ۱۹۵۰ به ندرت پیشرفتی در این زمینه دیده شد. با پیشرفت‌های عظیمی که در قابلیت‌های GPU و سایر منابع محاسباتی مرتبط انجام شد، این زمان حال حاضر بهترین زمان برای پیشرفت در تحقیقات یادگیری عمیق بوده است. اما این با یک هشدار همراه است؛ منابع محاسباتی بزرگتر و بهتر هم هزینه سنگینی دارند که ممکن است برای اکثر افراد و شرکت‌ها اقتصادی نباشد. بنابراین، در زمینه منابع به راحتی دسترسی‌پذیر، رویکرد سنتی به نظر می‌رسد که بهترین گزینه مقابل یادگیری عمیق است.

۷. نیاز به مجموعه داده‌های بزرگ و برچسب‌گذاری‌شده:

ما در زمانی زندگی می‌کنیم که در هر لحظه هزاران و هزاران پتابایت داده در سراسر جهان ایجاد و ذخیره می‌شود. این امر در ظاهر خبر خوبی است اما متأسفانه، بر خلاف باور متداول، ذخیره مقدار زیادی داده، به خصوص داده‌های تصویر، اقتصادی نه تنها نیست بلکه فرصت تجاری پایداری نیز ارائه نمی‌دهد. ممکن است شگفت‌زده شوید که بسیاری از شرکت‌ها دارای یک مجموعه داده غنی هستند. اما یا توانایی بهره‌مندی از آن را ندارند یا کسب‌وکار قانونی نمی‌تواند راه بیندازند. بنابراین، یافتن یک مجموعه داده مفید، برچسب‌گذاری شده و در سیاق آن کاربرد، وظیفه‌ای ساده برای یک راهکار یادگیری عمیق نیست.

۸. مدل‌های جعبه سیاه تفسیرناپذیر:

رویکردهای سنتی از روش‌های آماری قابل فهم و تفسیری مانند SVM و KNN برای یافتن ویژگی‌ها برای حل مشکلات متداول بینایی کامپیوتر استفاده می‌کنند. در مقابل، یادگیری عمیق شامل استفاده از لایه‌های پیچیده از شبکه‌های چندلایه پرسپترون (MLP) است. این MLPها ویژگی‌های اطلاعاتی را از تصاویر با فعال‌سازی مناطق مرتبط در تصاویر استخراج می‌کنند که اغلب قابل تفسیر نیستند. به عبارت دیگر، نمی‌توان به قطعیت گفت چرا بخش‌های خاصی از یک تصویر فعال شدند در حالی که بخش دیگر فعال نشد.

۹. کوچک و آسان برای حمل و یا استقرار درون یک میکروپردازنده:

به علاوه از اینکه محاسباتی سنگین هستند، مدل‌های مورد استفاده در یک رویکرد یادگیری عمیق به اندازه قابل توجهی بزرگتر از رویکردهای سنتی هستند. این مدل‌ها اغلب از اندازه چند صد مگابایت تا یک یا دو گیگابایت تغییر می‌کنند. در حالی که در مقابل، رویکردهای سنتی معمولاً یک مدل با اندازه چند مگابایتی تولید می‌کنند.

۱۰. دقت پیش‌بینی‌های دو رویکرد:

یکی از عوامل موفقیت یادگیری عمیق برجسته شدن نسبت به دستاوردهای رویکردهای سنتی، دقت بسیار بالای پیش‌بینی‌هاست. این یک پیشرفت عظیم در اوایل دهه ۹۰ تا اوایل قرن ۲۱ بود که یان لوکان و همکارانش با LeNet به وجود آمدند. این مدل از دقت‌های پیشین که با رویکردهای سنتی به دست آمده بودند، کاملاً فراتر رفت. از آن زمان، یادگیری عمیق تقریباً به عنوان ابزار معمول برای هر مسئله بینایی کامپیوتری در نظر گرفته می‌شود.

۳

استفاده از Max Pooling به جای Avg Pooling در شبکه‌های عصبی کانولوشنی (CNN) در بینایی ماشین به دلایل مختلف ترجیح داده می‌شود از جمله:

۱. حفظ ویژگی‌های برجسته:

Max Pooling با انتخاب بیشینه‌ی مقدار در هر ناحیه، ویژگی‌های برجسته و حساس به جزئیات را حفظ می‌کند. این باعث می‌شود که شبکه اطلاعات مهم را در طول لایه‌های پولینگ حفظ کند و از افت دقت در تشخیص ویژگی‌های مهم جلوگیری شود.

۲. مقاومت به نویز:

Max Pooling مقاومت بیشتری نسبت به نویز دارد. اگر در یک ناحیه از تصویر نویز وجود داشته باشد و این نویز تاثیر بیشینه‌گیری را داشته باشد، اثر آن بر کل ویژگی تاثیر زیادی نخواهد داشت. این موضوع به شبکه این امکان را می‌دهد که در مقابل نویزهای کوچک و تغییرات جزئی در تصاویر مقاومت نشان دهد.

۳. اهمیت نقاط کلیدی:

در بسیاری از حالات، نقاط کلیدی تصویر (مثل لبه‌ها و نقاط برجسته) اطلاعات مهمی در مورد شیء در تصویر حاصل می‌کنند. Max Pooling با انتخاب بیشینه‌ی مقدار، به ویژگی‌های کلیدی و برجسته اهمیت می‌دهد و این امکان را فراهم می‌کند که شبکه به ویژگی‌های مهم تر بیشتر دسترسی داشته باشد نه‌ای‌تا به تفسیر آسان‌تر دست بیابد.

۴. کاهش ابعاد:

یکی از اهداف مهم در طراحی شبکه‌های عصبی کانولوشنی، کاهش ابعاد تصاویر است. با افزایش تعداد لایه‌ها و افزودن عمق به شبکه، ابعاد تصاویر می‌توانند به سرعت افزایش یابند. Max Pooling به شبکه این امکان را می‌دهد که از تصاویر به صورت مکمل و با اندازه‌های کوچکتر بهره‌مند شود. به عنوان مثال، فرض کنید داریم تصویر ورودی با ابعاد 100×100 داریم. اگر از یک لایه Max Pooling با اندازه 2×2 استفاده کنیم، ابعاد تصویر به نصف (50×50) کاهش می‌یابد. این کاهش ابعاد کمک می‌کند تا تعداد پارامترها کاهش یابد و همچنین محاسبات سریع‌تر انجام شود. در مواردی که تصاویر بزرگ یا شبکه عصبی عظیم استفاده می‌شود، این کاهش ابعاد باعث می‌شود که آموزش شبکه سریع‌تر و موثرتر انجام شود. از این رو، Max Pooling به عنوان یک عنصر کلیدی در تسهیل آموزش و استفاده از شبکه‌های عصبی کانولوشنی شناخته می‌شود.

۵. عدم وابستگی به میانگین:

Avg Pooling ممکن است به عنوان یک میانگین‌گیری عمل کند که ممکن است باعث از دست رفتن اطلاعات مهم در نواحی کلیدی شود. Max Pooling این مشکل را حل می‌کند زیرا به جای میانگین، بیشینه‌ی مقدار را انتخاب می‌کند.

به طور کلی، Max Pooling بهتر مناسب برای استخراج ویژگی‌های برجسته و حفظ اطلاعات مهم در تصاویر است. با این حال، در برخی موارد خاص، ممکن است Avg Pooling مورد استفاده قرار گیرد، اما اکثراً در شبکه‌های عصبی کانولوشنی، Max Pooling به عنوان انتخاب اصلی در لایه‌های پولینگ به کار گرفته می‌شود.

۴

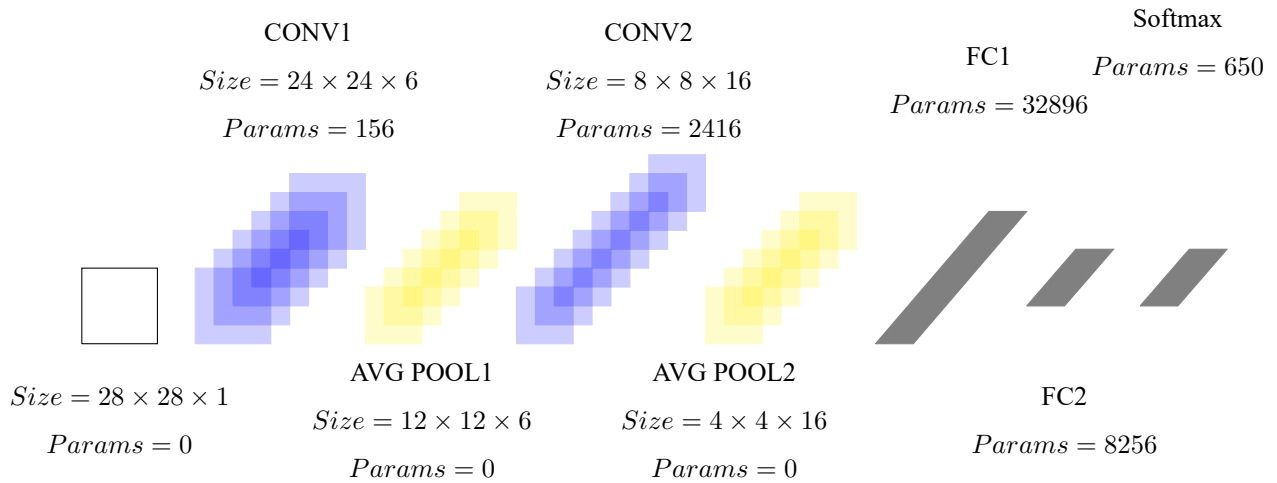


Figure 1

$$\text{size after applying convolution layer} = \left\lfloor \frac{n+2p-f}{s} + 1 \right\rfloor \times \left\lfloor \frac{n+2p-f}{s} + 1 \right\rfloor \times k$$

Input Layer:

- Size: $28 \times 28 \times 1$
- Params: 0
- Number of Channels: 1

CONV1 Layer:

- Filter size: 5×5
- Stride: 1
- Number of filters (k): 6
- Padding: 0
- Size: $\left\lfloor \frac{28+2 \times 0-5}{1} + 1 \right\rfloor \times \left\lfloor \frac{28+2 \times 0-5}{1} + 1 \right\rfloor \times 6 = 24 \times 24 \times 6$
- Params: $(5 \times 5 \times 1 + 1) \times 6 = 156$ (weights + bias for each filter)

Average Pooling Layer 1:

- Filter size: 2×2
- Stride: 2
- Size: $\left\lfloor \frac{24+2 \times 0-2}{2} + 1 \right\rfloor \times \left\lfloor \frac{24+2 \times 0-2}{2} + 1 \right\rfloor \times 6 = 12 \times 12 \times 6$
- Params: 0

CONV2 Layer:

- Filter size: 5×5
- Stride: 1
- Number of filters (k): 16
- Padding: 0
- Size: $\left\lfloor \frac{12+2 \times 0-5}{1} + 1 \right\rfloor \times \left\lfloor \frac{12+2 \times 0-5}{1} + 1 \right\rfloor \times 16 = 8 \times 8 \times 16$
- Params: $(5 \times 5 \times 6 + 1) \times 16 = 2416$ (weights + bias for each filter)

Average Pooling Layer 2:

- Filter size: 2×2
- Stride: 2
- Size: $\left\lfloor \frac{8+2 \times 0-2}{2} + 1 \right\rfloor \times \left\lfloor \frac{8+2 \times 0-2}{2} + 1 \right\rfloor \times 16 = 4 \times 4 \times 16$
- Params: 0

FC1 Layer (Fully Connected):

- Number of neurons: 128
- Params: $(4 \times 4 \times 16 + 1) \times 128 = 32896$ (weights + bias for each neuron)

FC2 Layer (Fully Connected):

- Number of neurons: 64
- Params: $(128 + 1) \times 64 = 8256$ (weights + bias for each neuron)

Softmax Layer:

- Number of neurons: 10 (assuming classification)
- Params: $(64 + 1) \times 10 = 650$ (weights + bias for each class)

$$\text{Sum Parameters} = 156 + 2416 + 32896 + 8256 + 650 = 44374$$

۵

۱. استفاده از Receptive Fields کوچک:

به جای استفاده از فیلترهای بزرگتر مانند 7×7 یا 11×11 با گام‌های طولی بزرگ، از فیلترهای 3×3 با گام ۱ استفاده می‌شود. این فیلترها در تمام شبکه به اندازه ۱ پیکسل در هر مرحله با تصویر ورودی همگام می‌شوند. با این رویکرد، از نظر شبکه‌های عصبی کانولوشنی (ConvNets) کوچک، هر دو لایه 3×3 در واقع دارای یک Receptive Fields پنج در پنج هستند. به عبارت دیگر، دو لایه 3×3 متوالی می‌توانند اطلاعات معادل یک لایه 5×5 را ادغام کنند. این امر به مدل این امکان را می‌دهد که از تصاویر با جزئیات بیشتری اطلاعات استخراج کند.

۲. زیاد کردن تعداد لایه‌ها و افزایش عمق:

استفاده از فیلترهای 3×3 به جای فیلترهای بزرگتر، این امکان را فراهم می‌کند که تعداد لایه‌ها در شبکه افزایش یابد. برای مثال، سه لایه 3×3 معادل با یک لایه 7×7 است. این افزایش در عمق باعث می‌شود که مدل قابلیت یادگیری و تعمیم بیشتری پیدا کند.

۳. کاهش تعداد پارامترها:

از دیدگاه تعداد پارامترها، سه لایه 3×3 کمترین تعداد پارامتر را دارند نسبت به یک لایه 7×7 . این امر به معنای کاهش تعداد وزن‌ها و پارامترها در شبکه است، که از یک سو به افت کمی در توانایی یادگیری منجر می‌شود، اما از سوی دیگر باعث می‌شود که مدل کم‌پیچیده‌تر و کارآمدتر باشد. فیلتر 3×3 تعداد پارامترهای کمتری نسبت به فیلترهای بزرگتر دارد. این موضوع باعث می‌شود که مدل‌ها سبک‌تر شوند و از نظر محاسباتی کارآمدتر باشند.

۴. افزایش ناحیه غیرخطی سازی:

با استفاده از سه لایه 3×3 متوالی به جای یک لایه 7×7 ، مدل شامل سه لایه تصحیح غیرخطی (non-linear rectification) می‌شود. این موضوع باعث افزایش توان تصمیم‌گیری مدل و افزایش تمایزپذیری تصمیم‌گیری می‌شود.

۵. قابلیت ترکیبی:

فیلتر 3×3 به دلیل اندازه کوچکتر، قابلیت ترکیبی بیشتری دارد. با ترکیب چندین فیلتر 3×3 ، می‌توان ویژگی‌های بزرگتر و پیچیده‌تری را استخراج کرد. به عبارت دیگر، تأثیرات یک فیلتر بزرگتر را می‌توان با استفاده از چندین فیلتر 3×3 شبیه‌سازی کرد.

۶. استفاده مکرر:

استفاده مکرر از فیلترهای 3×3 در لایه‌های متوالی، امکان آموزش سلسله‌مراتبی و افزایش عمق (depth) را بهبود می‌بخشد. این موضوع به مدل این امکان را می‌دهد که ویژگی‌های سطوح مختلف را استخراج کند

به طور کلی، استفاده از فیلترهای 3×3 در شبکه‌های عصبی کانولوشنی به دلایل فوق منجر به افزایش قابلیت یادگیری، کاهش تعداد پارامترها، و افزایش قابلیت تمایزپذیری می‌شود. این رویکرد توسط شبکه‌های عصبی معروفی نظیر GoogLeNet و شبکه‌های دیگر با موفقیت به کار گرفته شده است.

۶

ResNet نوعی از معماری‌های شبکه‌های عصبی پیچشی (CNN) است که residual connections را معرفی می‌کند، که مشکل گرادیان محو شونده را حل کرده و آموزش شبکه‌های بسیار عمیق را تسهیل می‌کند. در اکثر موارد، ResNet نسبت به معماری‌های سنتی CNN مزایای زیادی دارد. به ویژه زمانی که با شبکه‌های بسیار عمیق سروکار داریم. با این حال، مواقعی وجود دارد که استفاده از ResNet مزیت قابل توجهی ندارد یا حتی ممکن است مناسب نباشد:

۱. شبکه‌های کم عمق:

ResNet برای حل چالش‌های آموزش شبکه‌های بسیار عمیق طراحی شده است. اگر کار یا مجموعه داده شما نسبتاً ساده باشد و به یک معماری کم عمق نیاز نداشته باشد، استفاده از یک CNN ساده ممکن است کافی باشد و افزودن اتصالات باقی‌مانده به ارتقاء چشمگیری منجر نشود.

۲. داده محدود:

قابلیت ResNet در یادگیری ویژگی‌های پیچیده در مواردی که تعداد داده‌های آموزشی کم است، مفیدتر نیست. در شرایطی که مجموعه داده کوچک است، یک معماری ساده‌تر ممکن است کمتر دچار بیش‌برازش (overfitting) شده و بهتر عمل کند.

۳. منابع محاسباتی:

آموزش شبکه‌های عمیق، به ویژه با اتصالات باقی‌مانده، ممکن است مصرف منابع محاسباتی زیادی را به همراه داشته باشد. اگر محدودیت منابع محاسباتی دارید، استفاده از یک CNN ساده‌تر ممکن است عملی‌تر باشد.

۴. قابلیت تفسیر مدل:

اتصالات باقی‌مانده ممکن است تفسیر ویژگی‌های یادگرفته‌شده را دشوارتر کنند. در برخی موارد، به ویژه زمانی که قابلیت تفسیر برای شما مهم است، یک CNN سنتی با ساختار کم‌عمق‌تر ممکن است ترجیح داده شود.

۵. معماری‌های تخصصی:

برای وظایف یا حوزه‌های خاص، ممکن است معماری‌های خاص CNN وجود داشته باشد که مناسب‌تر باشند. به عنوان مثال، اگر وظیفه شما مرتبط با داده‌های متوالی باشد، شبکه‌های عصبی بازگشتی (RNN) یا شبکه‌های حافظه کوتاه‌مدت طولانی (LSTM) ممکن است مناسب‌تر باشند.

۶. انتقال یادگیری:

در مواردی که مدل‌های پیش‌آموزش داده‌شده برای معماری‌های خاص CNN موجود و وظیفه شما با ویژگی‌های یادگرفته‌شده توسط آن مدل همخوانی داشته باشد، ممکن است موثرتر باشد که از مدل پیش‌آموزش داده‌شده به جای ResNet استفاده کنید.

۷

کاربردهای لایه‌ی مانولوشنی 1×1 Conv (به شرح زیر است:

۱. کاهش/افزایش ابعاد:

1×1 Conv برای کاهش یا افزایش تعداد کانال‌ها در نقشه ویژگی استفاده می‌شود، که به طور مؤثر اعماق داده را در عین حفظ اطلاعات فضایی (ارتفاع و عرض) کاهش یا افزایش می‌دهد.

۲. کاهش بار محاسباتی: با استفاده از 1×1 Conv قبل از لایه‌های کانولوشن بزرگ‌تر (مثل 3×3 یا 5×5)، تعداد عملیات به طریق قابل توجهی کاهش می‌یابد که منجر به کارایی محاسباتی می‌شود. این به ویژه زمانی مفید است که با نقشه‌های ویژگی ورودی بزرگ سر و کار داریم.

۳. غیرخطیت بیشتر: 1×1 Conv غیرخطیت را به شبکه می‌آورد و با اعمال توابع فعال‌سازی به خروجی کانولوشن، به توانایی ابزار مدل کمک می‌کند.

۴. ایجاد شبکه‌های عمیق‌تر ("Bottle-Neck"): 1×1 Conv در طراحی "لایه Bottle-Neck" استفاده می‌شود، همانند معماری ResNet. این لایه به کاهش و بازیابی ابعاد در دنباله‌ای از لایه‌های کانولوشن کمک می‌کند و ترتیب آموزش شبکه‌های بسیار عمیق را تسهیل می‌دهد.

۵. ایجاد مدل‌های کوچک‌تر و با دقت بالاتر ("Fire module"): در مدل‌هایی مانند SqueezeNet، 1×1 Conv جزء حیاتی در "Fire module" برای کاهش کانال‌های ورودی قبل از تغذیه به لایه گسترشی استفاده می‌شود. این منجر به ایجاد یک مدل کوچک‌تر با تعداد کمتری پارامتر در حالی که دقت را حفظ می‌کند.

به طور خلاصه، 1×1 Conv ابزار چند منظوره‌ای در معماری شبکه‌های عصبی کانولوشنی است، که به منظور کاهش ابعاد، کارایی محاسباتی، معرفی غیرخطیت، ایجاد شبکه‌های عمیق‌تر، و ایجاد مدل‌های کوچک‌تر با دقت بالاتر مورد استفاده قرار می‌گیرد.

منابع

- [1] <https://medium.com/analytics-vidhya/talented-mr-1x1-comprehensive-look-at-1x1-convolution-in-deep-learning-f6b355825578>
- [2] <https://machinelearningmastery.com/introduction-to-1x1-convolutions-to-reduce-the-complexity-of-convolutional-neural-networks/>
- [3] <https://www.quora.com/Why-are-3x3-filters-the-standard-in-Convolutional-Neural-Networks-CNN-Is-there-any-advantage-over-2x2-or-4x4-filters-for-example>
- [4] <https://www.coursera.org/lecture/convolutional-neural-networks/networks-in-networks-and-1x1-convolutions-ZTb8x>