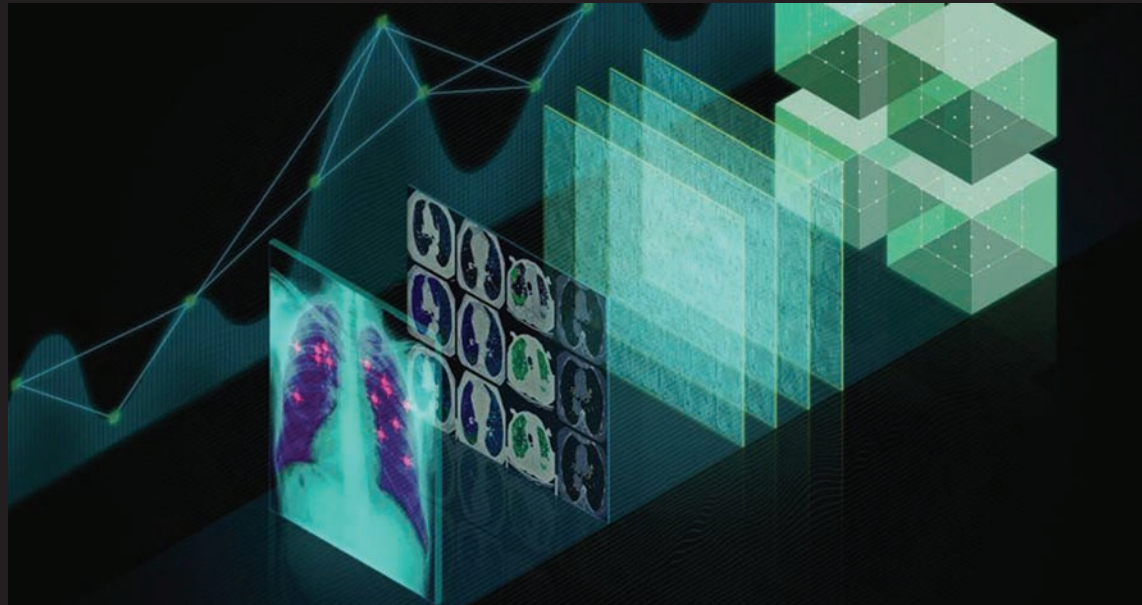




پردازش تصاویر پزشکی با استفاده از شبکه‌های عصبی کانولوشن عمیق

■ علی کارساز / استادیار / گروه مهندسی برق - کنترل / موسسه آموزش عالی خراسان / مشهد - ایران / karsaz@khorasan.ac.ir
 ■ سبورا محمدیان روشن / کارشناسی ارشد / گروه مهندسی برق و مهندسی پزشکی / موسسه آموزش عالی خراسان / مشهد - ایران / saboora.m.roshan@gmail.com

چکیده

حوزه پردازش تصاویر پزشکی بازه وسیعی از کاربردها از تشخیص دیابت چشمی از روی تصاویر شبکه چشم تا بخش‌بندی تصاویر MRI جهت تشخیص تومورهای مغز انسان را در بر می‌گیرد. نگرش‌های گوناگون دسته‌بندی و خوشه‌بندی مبتنی بر یادگیری ماشین در مقالات جهت بهبود دقت در غربالگری بیماری‌ها ارایه شده است. بعضی از این روش‌ها مبتنی بر استخراج ویژگی‌ها از روی تصاویر پزشکی به صورت دستی بوده و توسط متخصصان پردازش تصویر با صرف زمان و انرژی زیاد، صورت می‌پذیرد. در سال‌های اخیر روشی جدید برای تشخیص و طبقه‌بندی تصاویر پزشکی بدون نیاز به استخراج ویژگی‌ها به صورت دستی، مبتنی بر شبکه‌های عصبی کانولوشن ارایه شده است. این دسته از شبکه‌های عصبی که مبتنی بر فرآیند یادگیری عمیق ارایه شده‌اند نسبت به شبکه‌های عصبی معمولی به علت داشتن لایه‌های کانولوشن و مخفی بیشتر در زمینه کار با ورودی‌های با ابعاد بالاتر مانند تصاویر دارای توانمندی بیشتری هستند. یک معضل جدی در آموزش یک شبکه عصبی کانولوشن با یادگیری عمیق، به موضوع آموزش آنها از ابتدا باز می‌گردد این معضل ناشی از کمبود داده‌های طبقه‌بندی شده جهت آموزش و زمان‌بر بودن فرآیند آموزش تا یک همگرایی مناسب است. بنابراین یک روش متداول برای آموزش شبکه‌های عصبی کانولوشن بر روی داده‌های پزشکی، بر اساس بازتنظیم شبکه‌های از پیش آموزش یافته، می‌باشد. برخی از مهم‌ترین و نیرومندترین شبکه‌های عصبی کانولوشن که بر روی بانک‌های اطلاعات تصاویر، نظیر بانک اطلاعات تصاویر ImageNet با بیش از یک میلیون تصویر، آموزش دیده‌اند شامل شبکه سیفرانت (CifarNet)، الکسنت (AlexNet) و گوگلنت (GoogleNet) است. که جهت بازتنظیم آنها در تشخیص تصاویر پزشکی با کاربردهای خاص و با تعداد محدودی از تصاویر، می‌توان بهره برد.

کلمات کلیدی: شبکه‌های عصبی کانولوشن، شبکه کانولوشن گوگلنت، شبکه الکسنت، شبکه سیفرانت، بانک اطلاعات تصاویر ImageNet.

Medical image processing using deep convolutional neural networks

■ Ali Karsaz / Assistant Professor, Electrical and Bioelectric Engineering Department, Khorasan Institute of Higher Education, Mashhad, Iran
 karsaz@khorasan.ac.ir
 ■ Saboora Mohammadian Roshan/ M.Sc. Electrical and Bioelectric Engineering Department, Khorasan Institute of Higher Education, Mashhad, Iran/ saboora.m.roshan@gmail.com

Abstract:

The field of medical image processing includes a wide range of applications from automated screening of diabetic retinopathy based on retinal images to MRI segmentation for tumor recognition. Various machine learning classification and clustering approaches have been studied in literature with the purpose of improving the accuracy of the screening methods. Some studies used manually feature extraction of fundus images by image processing experts. In recent years, a new approach for image classification and diagnosis without using any manual feature extraction is proposed based on convolutional neural network (CNN). The CNNs are based on deep learning concept have more convolutional and hidden layers and are more powerful involving the high dimension inputs such as medical images. In medical imaging and diagnosis, training a deep CNN from scratch is difficult because it requires a large amount of labeled training data and the training procedure is a time consuming task to ensure proper convergence. Therefore, a very common method to train CNNs for medical diagnosis is fine-tuning a pre-trained CNN. Some of these powerful pre-trained CNNs are the GoogleNet, CifarNet and AlexNet which have been trained on the ImageNet as a large databas

Keywords: Diabetic retinopathy; Convolutional neural network (CNN), GoogleNet, Kaggle retinopathy database.

مهم‌ترین مزیت استفاده از شبکه‌های عصبی کانولوشن^۱ (CNN) توانایی آن در استخراج ویژگی‌های تصاویر به صورت خودکار با استفاده از مفهوم یادگیری عمیق^۲ است [۲، ۱]. به دلیل همین مزیت بسیار مهم در سال‌های اخیر، استفاده از CNN‌ها برای تشخیص بیماری‌ها در کاربردهای متفاوت پزشکی مورد توجه محققان قرار گرفته است. برای مثال، در زمینه تشخیص رتینوپاتی دیابتی (DR) یا دیابت چشمی که به علت بیماری دیابت در این دسته از بیماران، رگ‌های شبکیه چشم تغییر شکل داده و برای خون‌رسانی بهتر گشادتر می‌شوند، از شبکه‌های CNN، جهت استخراج ویژگی‌های تصاویر شبکیه کمک گرفته شده و سپس به روش یادگیری گروهی، رگ‌های خونی شبکیه چشم بخش‌بندی شده‌اند [۳، ۴]. از CNN‌ها برای تشخیص مراحل پیشرفت بیماری DR نیز استفاده می‌شود [۵، ۶]. همچنین در یک مطالعه دیگر، مقایسه‌ای بین دو ساختار CNN با ابعاد و تعداد لایه‌های چندگانه برای تشخیص هرچه دقیق‌تر DR انجام شده است [۸-۱۰]. از دیگر کاربردهای بکارگیری CNN‌ها در تشخیص و دسته‌بندی می‌توان به تشخیص صحیح پولیپ‌ها^۳ [۱۱] در طول فیلم‌برداری کلونوسکوپی^۴ [۱۲]، تشخیص اتومات رگ‌گرفتگی‌های ریوی^۵ در تصاویر سی‌تی (CT) اسکن^۶ [۱۳]، تشخیص به کمک کامپیوتر تقسیم سلولی^۷ سرطان سینه در بانک‌های اطلاعات پاتولوژی [۱۴]، تشخیص غدد لنفاوی^۸ در تصاویر CT [۱۵]، تشخیص غدد لنفاوی شکمی^۹ [۱۶]، دسته‌بندی بیماری‌های منافی ریوی^{۱۰} [۱۷] و تشخیص اتوماتیک آناتومی در تصاویر CT [۱۸] نام برد. کاربردهای CNN‌ها تنها به موضوع سیستم‌های تشخیصی محدود نمی‌شود بلکه به تازگی در زمینه اندازه‌گیری و بخش‌بندی تصاویر حوزه پزشکی نیز دارای عملکرد مناسبی بوده‌اند. برخی از این کاربردها شامل بخش‌بندی لوزالمعده^{۱۱} در اسکن توموگرافی [۱۸]، [۱۹]، بخش‌بندی تصاویر مغز کودک^{۱۲} [۲۰]، بخش‌بندی غشا عصبی در تصاویر الکترومیکروسکوپی [۲۱]، بخش‌بندی تصاویر غضروف زانو^{۱۳} در اسکن MRI [۲۲] و سنجش ضخامت لایه‌های عروق^{۱۴} [۲۳] و ده‌ها زمینه دیگر بوده است.

در این مقاله در بخش دوم به بررسی ساختار نوعی یک شبکه عصبی کانولوشن و معرفی لایه‌های مختلف آن پرداخته می‌شود. در بخش سوم، چند نمونه از معروف‌ترین شبکه‌های عصبی کانولوشن از قبیل شبکه سیفانت^{۱۵}، الکسنت^{۱۶} و گوگلنت^{۱۷} معرفی و بررسی می‌شوند. در نهایت در بخش چهارم نتیجه‌گیری و جمع‌بندی ارائه می‌گردد.

۲- شبکه‌های عصبی کانولوشن

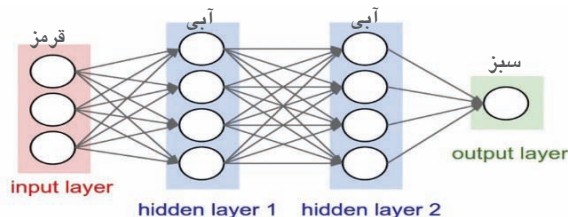
شبکه‌های عصبی کانولوشن (CNN) مدلی از شبکه‌های عصبی مصنوعی بوده که همانند شبکه‌های عصبی معمولی^{۱۸} (TNN) از نورون‌ها، لایه‌ها و وزن‌ها تشکیل شده است. مهم‌ترین تفاوت CNN با TNN‌ها، به قابلیت ویژه CNN‌ها در یادگیری‌های عمیق مانند تشخیص تصاویر، صداها و دست‌خط‌های متفاوت باز می‌گردد. بنابراین ورودی این شبکه‌ها می‌تواند شامل ماتریس‌هایی از اطلاعات با ابعاد بالاتر باشد. مطالعه‌ای که در سال ۱۹۶۸ میلادی صورت گرفت نشان داد که قشر بینایی مغز برای پردازش اطلاعات تصاویر از الگوی پیچیده‌ای استفاده می‌نماید [۲۴]. نواحی ادراکی که قشر بینایی در آن قرار دارد، همانند فیلترهای محلی بر روی اطلاعات تصویر اعمال می‌شود سلول‌های ساده‌تر برای تشخیص ویژگی‌های سطح پایین‌تر در نواحی ادراکی همچون لبه‌ها کاربرد دارند، همچنین سلول‌های پیچیده قادر به تشخیص ویژگی‌های مهم‌تر و اختصاصی‌تر و در سطوح

بالاتر می‌باشند. تشخیص ویژگی‌های اختصاصی‌تر نتیجه و ترکیبی از ویژگی‌های سطح پایین می‌باشند. این عملکرد مغز الهام‌بخش شبکه‌های عمیق عصبی امروزی شده است. مفهوم شبکه کانولوشن برای نخستین بار در سال ۱۹۸۰ توسط فکوشیما^{۱۹} مطرح گردید [۲۵]. در سال بعد شبکه‌های کانولوشن امروزی به منظور استفاده در شبکه‌های عصبی بررسی شدند. اما به دلیل نیاز به سخت‌افزارها و پردازشگرهای گرافیکی قوی استفاده از این شبکه برای تشخیص تصاویر تا سال ۲۰۱۲ که به شکل اختصاصی برای تشخیص تصاویر ارایه و معروف گردید به تعویق افتاد [۲۶].

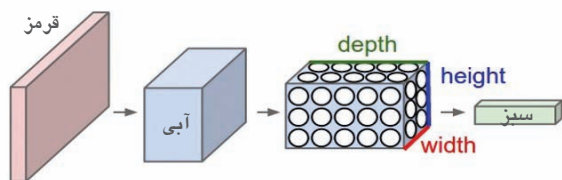
۲-۱- ساختار شبکه عصبی کانولوشن

شبکه عصبی کانولوشن مدلی از شبکه‌های عصبی چندلایه هستند که به طور معمول شامل لایه‌های کانولوشن، پولینگ^{۲۰} و تمام متصل^{۲۱} (FC) می‌باشند [۲۶]. در یک TNN، شبکه عصبی، ورودی را دریافت و آن را به تعدادی لایه مخفی می‌فرستد. هر لایه مخفی از تعدادی نورون به شکل یک بردار تشکیل شده است که هر نورون به طور کامل به تمامی نورون‌های لایه قبلی خود متصل می‌باشد نورون‌ها در هر لایه به طور مستقل عمل می‌کنند و در ارتباطات و وزن‌ها با یکدیگر شریک نمی‌شوند. لایه آخر این شبکه، لایه خروجی بوده که مسوول نهایی تعیین کلاس داده‌ها است. اما در CNN‌ها، ورودی یک ماتریس چند لایه یا دارای عمق مانند یک تصویر است و لایه‌های این شبکه بر خلاف TNN‌ها، متشکل از نورون‌هایی در سه بعد هستند. که در برخی از مقالات به آنها تانسور نیز گفته می‌شود.

در این شبکه‌ها نورون‌های یک لایه تنها به ناحیه کوچکی از لایه قبلی متصل می‌باشند و به تمامی نورون‌های لایه قبلی متصل نیستند. شکل (۱) تفاوت شبکه عصبی کانولوشن با شبکه عصبی معمولی را نشان می‌دهد. همان‌طور که در شکل مشخص است تمامی نورون‌ها در یک TNN به هم متصل می‌باشند اما در یک CNN نورون‌ها در سه بعد قرار گرفته‌اند. در شکل (۱-ب)، لایه‌ی ورودی که با رنگ قرمز مشخص شده است می‌تواند یک تصویر ورودی به شبکه باشد که دارای طول، عرض و عمق بوده و برای تصاویر با فرمت RGB معرف سه کانال قرمز، سبز و آبی در تصویر است [۲۷].



(الف)



(ب)

شکل ۱: مقایسه شبکه عصبی معمولی (TNN) و شبکه عصبی کانولوشن (CNN)، (الف) شبکه عصبی معمولی، (ب) شبکه عصبی کانولوشن

۲-۱-۱- لایه کانولوشن

این لایه به عنوان اصلی‌ترین لایه در این دسته از شبکه‌های عصبی محسوب شده به همین دلیل، آنها را به عنوان شبکه‌های عصبی

ورودی با ابعاد $W_1 \times H_1 \times D_1$ به صورت روابط (۲)، (۳) و (۴) به دست می آید:

$$W_2 = \frac{W_1 - F + 2P}{S + 1} \quad (2)$$

$$H_2 = \frac{H_1 - F + 2P}{S + 1} \quad (3)$$

$$D_2 = K \quad (4)$$

در این روابط F ، P ، S و K به ترتیب نشان دهنده اندازه کرنل، مقدار zero padding، اندازه ی گام و تعداد فیلترها می باشد. از این روابط مشخص است که به ازای هر فیلتر به مقدار $F \times F \times D_1$ وزن داریم و با توجه به تعداد K فیلتر موجود، در مجموع، تعداد $(F \times F \times D_1) \times K$ وزن و K بایاس ایجاد می شود. بنابراین تعداد پارامترهایی که شبکه در یک لایه کانولوشن خود می باید آموزش ببیند، بسیار زیاد هستند.

۲-۱-۲- لایه پولینگ^{۲۹}

در یک شبکه عصبی کانولوشن به طور معمول پس از هر لایه کانولوشن یک لایه پولینگ قرار می گیرد. این لایه از آن جهت اهمیت دارد که باعث کاهش تعداد پارامترهایی می شود که بایستی آموزش ببینند. بنابراین با بکارگیری این لایه ضمن کاهش محاسبات مورد نیاز در بخش آموزش، باعث کنترل فرایادگیری^{۳۰} احتمالی در شبکه نیز می گردد. این لایه بر روی هر برش از لایه های عمق ورودی اعمال می شود و اندازه آن را تغییر می دهد. دو تابع عملکردی معروف این لایه، پولینگ بیشینه^{۳۱} و میانگین^{۳۲} نام دارند، که مدل نخست دارای کاربرد بیشتری در CNNها است. طبقه عمل پولینگ بیشینه بدین صورت است که در هر پنجره، بزرگترین پیکسل را به خروجی می فرستد. این پنجره بر روی تصویر همانند تابع کانولوشن از چپ به راست و از بالا به پایین با اندازه گام های مشخص شده حرکت می کند و نتیجه را به خروجی می فرستد. به دلیل اینکه این عملیات بر روی تمامی برش ها اعمال می گردد، عمق خروجی همان عمق ورودی به لایه پولینگ می باشد و تغییری نمی کند. رابطه مربوط به ماتریس خروجی با ابعاد $W_3 \times H_3 \times D_3$ که از ماتریس ورودی مرحله کانولوشن با ابعاد $W_2 \times H_2 \times D_2$ بدست می آید به صورت روابط زیر می باشد.

$$W_3 = \frac{W_2 - F_p}{S_p + 1} \quad (5)$$

$$H_3 = \frac{H_2 - F_p}{S_p + 1} \quad (6)$$

$$D_3 = D_2 \quad (7)$$

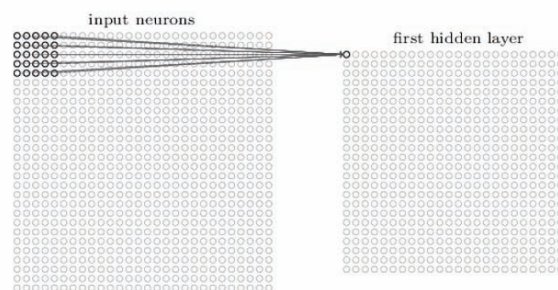
در روابط بالا، F_p و S_p به ترتیب اندازه ناحیه ادراکی یا کرنل و گام لایه پولینگ می باشند. شکل (۳) عملکرد لایه پولینگ بیشینه را نشان می دهد.

۲-۱-۳- لایه به طور کامل متصل

لایه به طور کامل متصل، لایه آخر یک شبکه عصبی کانولوشن محسوب می شود و مشابه شبکه عصبی معمولی اتصالات کاملی با خروجی لایه قبلی ایجاد می کند. این لایه ورودی را دریافت و سپس خروجی را به صورت برداری با N مولفه تولید می کند. تعداد کلاس هایی که شبکه مورد نظر آن را طبقه بندی می کند برابر با N است. در واقع یک شبکه CNN جهت تولید یک بردار خروجی با

کانولوشن می شناسند. وظیفه این لایه استخراج ویژگی ها است. این لایه عملیات کانولوشن را بر روی داده ورودی اعمال می کند و خروجی هایی به نام نقشه ویژگی^{۳۳} از این لایه بدست می آید. در نتیجه تمامی نورون ها در یک نقشه ویژگی، مجموعه ای از وزن ها و بایاس های مشابه و مشترکی دارند که باعث می شود، ویژگی های تصویر در موقعیت های مختلف قابل شناسایی باشند. از طرف دیگر این اشتراک وزن ها باعث کاهش تعداد پارامترهای مورد نیاز برای آموزش می گردد.

در شبکه های کانولوشن اتصالات به صورت نواحی کوچک و محلی صورت می گیرد. به بیان دیگر هر نورون در نخستین لایه مخفی به ناحیه کوچکی از نورون های ورودی متصل است. برای مثال این ناحیه می تواند 5×5 باشد. این ناحیه کوچک ۲۵ پیکسلی ناحیه ادراک محلی^{۳۴} یا کرنل کانولوشن^{۳۴} نامیده می شود [۲۸]. برای توضیح بیشتر می توان شکل (۲)، را با یک تصویر ورودی 28×28 در نظر گرفت. پنجره کرنل 5×5 بر روی پیکسل های ورودی از چپ به راست حرکت می کند هر پنجره به نورونی در لایه مخفی متصل می شود. بنابراین همان طور که در شکل (۲) مشخص است لایه مخفی شامل یک شبکه 24×24 نورونی خواهد بود.



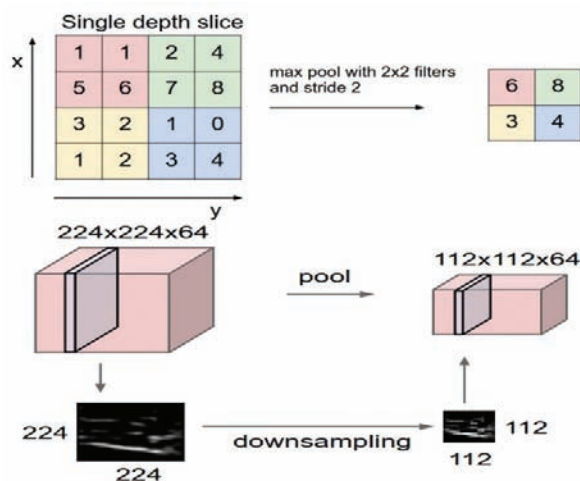
شکل ۲: عملیات کانولوشن در یک CNN با یک کرنل 5×5 . راست: نورون های لایه پنهان، چپ: نورون های لایه ورودی

در شکل (۲) هر نورون لایه مخفی دارای یک بایاس و تعداد 5×5 وزن می باشد که به ناحیه ادراکی خود متصل شده است. تمامی نورون های لایه مخفی مذکور که دارای ابعاد 24×24 خواهد بود، دارای وزن ها و بایاس مشترکی می باشند. به بیان دیگر خروجی نورون لایه کانولوشن $y_{w,h,m}$ در طول و عرض w ، h و عمق m به صورت رابطه (۱) قابل بیان است.

$$y_{w,h,m} = \quad (1)$$

$$f(\sum_{i=(w-1)S+1}^{(w-1)S+K} \sum_{j=(h-1)S+1}^{(h-1)S+K} \sum_{k=1}^N W_{k,m} x_{i,j,k} + b_m)$$

که در این رابطه f تابع فعالیت^{۳۵}، b_m بایاس مشترک نورون ها، $W_{k,m}$ وزن های 5×5 مشترک نورون ها و نهایتاً $x_{i,j,k}$ ورودی در موقعیت i, j, k و k می باشد. بکارگیری توابع فعالیت در لایه کانولوشن، باعث افزایش خصوصیات غیر خطی در خروجی می شوند. توابع فعالیت انواع متفاوتی دارند. چند نمونه از معروف ترین توابع فعالیت، تابع سیگموئید^{۳۶}، تابع تانژانت هایپربولیک (tanh)، تابع یکسوساز خطی^{۳۷} (ReLU)، تابع خطی نمای^{۳۸} (ELU) و Softmax است. بنابراین تمامی نورون های واقع در لایه مخفی اول به طور دقیق ویژگی مشابهی را در نواحی مختلف تصویر شناسایی می نمایند. در نهایت خروجی لایه ورودی یا نورون های لایه مخفی به نام نقشه ویژگی شناخته می شوند. ابعاد مربوط به ماتریس خروجی لایه کانولوشن $W_2 \times H_2 \times D_2$ که از ماتریس



شکل ۳: عملکرد لایه پولینگ با کرنل 2×2 جهت کاهش پارامترهای مورد آموزش شبکه

وزن جهت آموزش است. علاوه بر این آموزش شبکه از ابتدا به سخت‌افزارهای قوی همانند پردازشگرهای قوی گرافیکی نیاز دارد که در عمل با محدودیت سخت‌افزاری بیان شده آموزش این شبکه‌ها از ابتدا ناممکن می‌باشد [۲۹]. بنابراین به دو روش استخراج ویژگی و یا استفاده از روش بازتنظیم یک شبکه از پیش آموزش دیده در لایه‌های انتهایی آن می‌توان عمل نمود. در روش استخراج ویژگی، به کمک یک CNN، ویژگی‌های تصاویر پزشکی بدست آمده و به کمک یک روش طبقه‌بندی، همانند ماشین بردار پشتیبان می‌توان به دسته‌بندی مورد نظر دست یافت. این شیوه برای زمانی که تعداد نمونه‌های بسیار کمی موجود است دارای کاربرد بوده در این حالت بایستی ابتدا ویژگی‌های تصاویر را به کمک لایه‌های کانولوشن و پولینگ استخراج نموده و در نهایت این ویژگی‌ها را به‌عنوان ورودی به الگوریتم‌های طبقه‌بندی داد، تا طبقه‌بندی نهایی صورت گیرد. اما زمانی که تعداد نمونه‌ها بیشتر است ولی به اندازه‌ای نیست که بتوان کل پارامترهای شبکه را از ابتدا آموزش داد، و این نمونه‌ها با داده‌های آموزش شبکه اصلی بسیار متفاوت باشند بایستی دوباره از تنظیم یا بازتنظیم شبکه استفاده شود [۲۹]. در روش بازتنظیم یک CNN، ابتدا یک شبکه آموزش دیده که

N مولفه عددی طراحی می‌شود که هر عدد در این بردار خروجی، درصد احتمال تعلق به کلاس مورد نظر را نشان می‌دهد. بالاترین درصد بیان‌گر نتیجه نهایی آموزش شبکه است [۲۶]. در الگوریتم‌های یادگیری، بحث مهم در نظر گرفتن رابطه‌ای برای بدست آوردن ارتباط بین کلاس داده مورد نظر و فرض بدست آمده از نتیجه آموزش الگوریتم می‌باشد. این رابطه که با نام تابع تلفات نیز از آن یاد می‌شود بایستی با روش‌های مناسب، بهینه شود. بنابراین ایده اساسی در آموزش یک شبکه و یا الگوریتم یادگیری، بهینه کردن تابع تلفات با پیمودن گام‌های مناسب برای رسیدن به نتیجه بهتر است. روش‌های مهم بهینه‌سازی برای یک CNN شامل: گرادینت نزولی، گرادینت نزولی batch، گرادینت نزولی تصادفی و رادیان نزولی mini-batch، الگوریتم Adagrad، الگوریتم RMSprop و Adam می‌باشد. در شبکه‌های عصبی کانولوشن به دلیل اینکه تعداد پارامترهایی که آموزش می‌بینند بسیار زیاد است و همچنین این دسته از شبکه‌ها دارای لایه‌های عمیق هستند بنابراین در طی زمان آموزش، شبکه مستعد فرایادگیری می‌شود. این به این معنا است که شبکه برای داده‌های آموزش عملکرد خوبی از خود نشان می‌دهد اما این عملکرد با وارد شدن داده‌های تست دچار مشکل می‌شود. برای حل نمودن این مشکل راه‌های متفاوت بسیاری پیشنهاد شده است. یکی از روش‌های معروف، روش افزایش تعداد داده‌های آموزش و روش دیگر، روش اضافه نمودن لایه dropout مناسب در لایه به‌طور کامل متصل برای رفع این معضل است. روش افزایش داده‌ها در موضوع فرآیندهای یادگیری، یک روش مرسوم بوده که در آن با اعمال تغییراتی در تصاویر مانند چرخاندن، بزرگ‌نمایی، جابجایی و غیره به افزایش تعداد تصاویر ورودی به شبکه می‌پردازد. در روش dropout قبل از هر مرحله آموزش، تعدادی از نورون‌ها به‌صورت رندم و با نسبت تعیین‌شده در برنامه مورد نظر از فرآیند آموزش خارج می‌شوند و تنها نورون‌های باقی‌مانده در آموزش شرکت می‌کنند [۲۸].

۲-۲- باز تنظیم شبکه‌های از قبل آموزش دیده

در زمینه‌های پزشکی به علت کمبود تعداد تصاویر دارای طبقه‌بندی مناسب، آموزش کامل یک شبکه عصبی کانولوشن بسیار دشوار و یا غیرممکن است. به‌عنوان مثال در بانک اطلاعات چشم پزشکی کگل تنها حدود ۳۵ هزار تصویر شبکه چشم وجود دارد حال آنکه یک شبکه عصبی کانولوشن نظیر آکسنت دارای ۶۰ میلیون

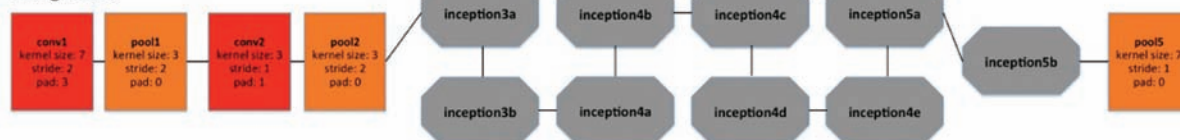
CifarNet



AlexNet



GoogLeNet

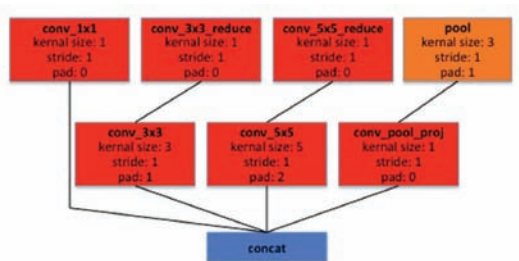


شکل ۴: نمایش ساده‌ای از ساختار سه شبکه عصبی کانولوشن معروف سیفانت، آکسنت و گوگلنت [۱۶]

۵ لایه کانولوشن، سه لایه پولینگ و دو لایه تماماً متصل بوده و حدود ۶۰ میلیون پارامتر آزاد جهت آموزش دارد.

۳-۳- شبکه CNN گوگل نت و مدل Inception-v3

شبکه عصبی کانولوشن گوگل نت به عنوان عمیق ترین شبکه CNN در بین مدل های رایج بوده مهمترین وجه تمایز این شبکه نسبت به مدل های قبلی مانند الکس نت یا سیفانت استفاده از ماژول Inception در ساختار خود است [۱]. به شکل کلی یک شبکه کانولوشن گوگل نت دارای دو لایه کانولوشن، دو لایه پولینگ و ۹ لایه Inception بوده هر لایه Inception خود دارای ۶ لایه کانولوشن و یک لایه پولینگ است. شکل (۷) ساختار یک گوگل نت را نشان می دهد. مدل Inception-v3 به عنوان یک مدل پیشرفته از شبکه عصبی کانولوشن گوگل نت، اولین بار توسط Szegedy و همکاران در سال ۲۰۱۵ معرفی گردید [۲]. در این مدل، کانولوشن های 1×1 ، 3×3 و 5×5 در یک ماژول یکسان و بر روی یک ورودی اعمال می شوند و خروجی تمامی این فرآیندها با یکدیگر الحاق شده تا به عنوان ورودی لایه بعدی در نظر گرفته شوند. شکل (۸) یک ماژول Inception از شبکه کانولوشن گوگل نت را نشان می دهد. به دلی ل الحاق مذکور نام دیگر ماژول Inception، استخراج کننده ی ویژگی چند لایه می باشد.



شکل ۸: یک ماژول Inception-a3 از شبکه عصبی کانولوشن گوگل نت

۴- نتیجه گیری

شبکه های عصبی مصنوعی کانولوشن در طول کمتر از یک دهه گذشته با توسعه قدرت پردازش کامپیوترها، در حوزه یادگیری عمیق سهم به سزایی را ایفا نموده به گونه ای که در کاربردهای مهم و حیاتی مانند پردازش تصاویر پزشکی و سیستم های اتومبیل خودران با موفقیت های روزافزونی همراه بوده اند.

بکارگیری شبکه های کانولوشن از قبیل سیفانت، گوگل نت و الکس نت به علت استفاده از لایه های عمیق بیشتر در مقایسه با شبکه های عصبی متداول، می تواند در پردازش تصاویر پزشکی و مقوله غربالگری بیماری ها تحول چشم گیری را ایجاد نماید.

پی نوشت ها

- 1 Convolutional neural network
- 2 Deep learning
- 3 Polyp type
- 4 Colonoscopy videos
- 5 Pulmonary embolism (PE)
- 6 CT scan
- 7 Mitotic cells
- 8 Lymph nodes
- 9 Abdominal lymph nodes
- 10 Interstitial lung diseases
- 11 Pancreas segmentation
- 12 Infant brain image segmentation
- 13 Knee cartilage segmentation
- 14 Carotid intima-media thickness measurement

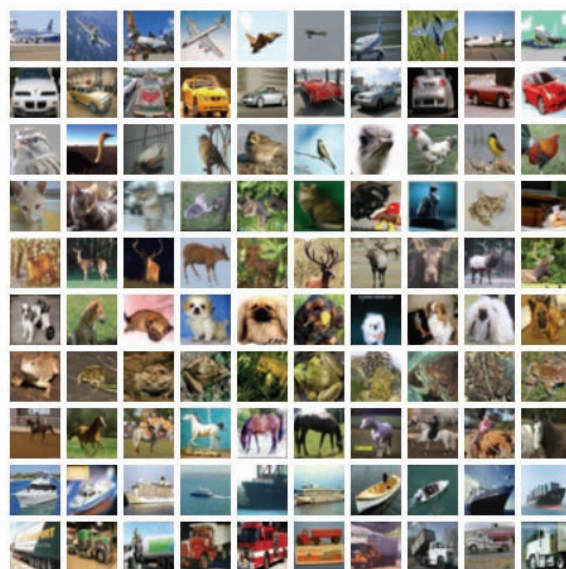
نتایج بسیار خوبی در بانک های تصاویر مطرح بدست آورده انتخاب می گردد، سپس لایه تماماً متصل نهایی با لایه تماماً متصل مورد نظر که قابلیت طبقه بندی به تعداد کلاس های مورد نظر را دارد، جایگزین می شود. در این فرآیند وزن های لایه های اولیه شبکه اصلی بدون تغییر وارد شبکه مورد نظر جدید می شوند و لایه های نهایی شبکه دوباره آموزش می بینند تا شبکه مورد نظر قابلیت طبقه بندی تصاویر جدید را داشته باشد. برخی از شبکه های عصبی کانولوشن معروف که در چند سال اخیر معرفی شده اند در ادامه مورد بررسی قرار می گیرد.

۳- انواع شبکه های عصبی کانولوشن

سه نمونه از پرکاربردترین CNN های موجود به خصوص در حوزه پردازش تصاویر پزشکی در این بخش مورد بررسی اجمالی قرار می گیرند. این سه شبکه عصبی کانولوشن شامل شبکه های سیفانت، الکس نت و گوگل نت بوده که جهت بازتنظیم آنها در تشخیص بیماری ها و یا بخش بندی های تصاویر پزشکی می توان بهره برد [۱۶]. شکل (۴) ساختار کلی این سه CNN را نشان می دهد.

۳-۱- شبکه CNN سیفانت

سیفانت اولین بار در دانشگاه تورنتو و در قالب تز دکتری در سال ۲۰۰۹ برای تشخیص اشیای بانک اطلاعات تصاویر سیفار ۱۰ معرفی شد [۳۰]. این بانک اطلاعات تصاویر شامل ۶۰۰۰۰ عکس با ابعاد 32×32 و در ۱۰ کلاس مختلف بوده لذا به ازای هر کلاس ۶۰۰۰ عکس متفاوت کوچک وجود دارد. شکل (۵) نمونه ای از تصاویر این بانک اطلاعاتی را نشان می دهد. سیفانت دارای سه لایه کانولوشن، سه لایه پولینگ و یک لایه تماماً متصل می باشد.



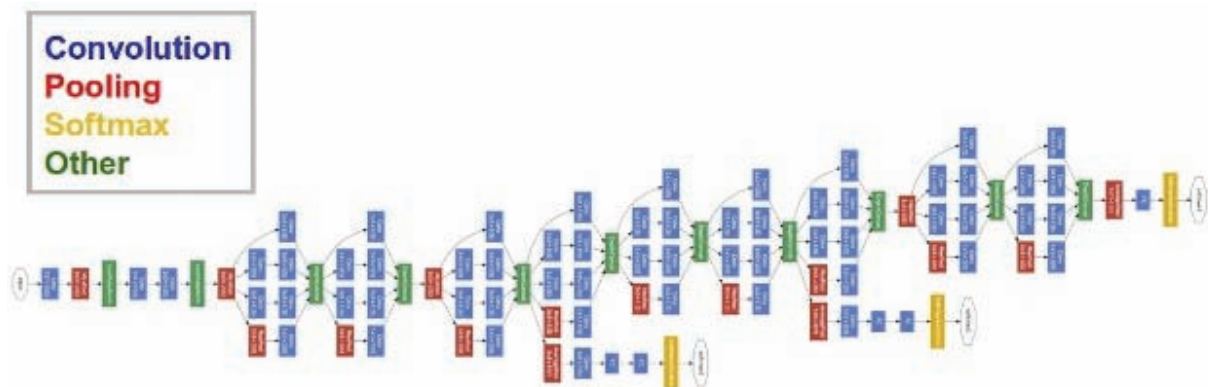
شکل ۵: چند نمونه از تصاویر بانک اطلاعاتی سیفار ۱۰، شامل ۶۰۰۰۰ عکس در ۱۰ کلاس متفاوت با ابعاد 32×32

۳-۲- شبکه CNN الکس نت

این شبکه در ۲۶ امین کنفرانس سیستم های پردازش اطلاعات عصبی در سال ۲۰۱۲ معرفی گردید [۲۶]. این شبکه برتری خود را بر شبکه های عصبی غیر عمیق برای تشخیص تصاویر در مسابقه معروف تشخیص تصاویر ابعاد بزرگ ImageNet2 به اثبات رساند [۳۱]. بانک اطلاعات ImageNet شامل $1/2$ میلیون تصویر با ابعاد 256×256 بوده که در ۱۰۰۰ کلاس مختلف گردآوری شده اند. شکل (۶) چند نمونه از تصاویر این بانک اطلاعات تصاویر را نشان می دهد. الکس نت شامل



شکل ۶: چند نمونه از تصاویر بانک اطلاعات تصاویر ImageNet. شامل ۱/۲ میلیون تصویر مختلف در ۱۰۰۰ کلاس متفاوت



شکل ۷: مدل شبکه عصبی کانولوشن گوگلنت با ماژولهای Inception به عنوان عمیق ترین CNN موجود

- 36 Support vector machine
- 37 Cifar10 database
- 38 Neural Information Processing Systems Conference

مراجع

- [1] C. Szegedy et al., "Going deeper with convolutions," IEEE Conference on computer Vision and Pattern Recognition (CVPR), 2015.
- [2] C. Szegedy et al., "Rethinking the Inception Architecture for Computer Vision," IEEE Conference on Computer Vision, Dec. 2015.
- [3] Y. Wang, G. Cao, B. Wei and G. Yang, "Hierarchical retinal blood vessel segmentation based on feature and ensemble learning," J. Neurocomput., vol. 149, pp. 708–717, Feb. 2015.
- [4] F. C. Harry Pratta, Deborah M. Broadbent, P. Simon P. Harding and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," International Conference On Medical Imaging Understanding and Analysis, July 2016.
- [5] H. H. Vo and A. Verma, "New deep neural nets for fine-grained diabetic retinopathy recognition on hybrid color space," IEEE Int. Symposium on Multimedia, Dec. 2016.
- [6] Y. M. S. Reddy, R. E. Ravindran and K. H. Kishore, "Diabetic retinopathy through retinal image analysis: A review," Int. J. of Engineering

- 15 CifarNet
- 16 AlexNet
- 17 GoogleNet
- 18 Traditional neural network
- 19 Fukushima
- 20 pooling layer
- 21 Fully connected layer
- 22 Feature map
- 23 Local receptive field
- 24 Convolutional kernel
- 25 Activation function
- 26 Sigmoid
- 27 Rectified linear unit
- 28 Exponential linear unit
- 29 Over fitting
- 30 Max-pooling
- 31 Average-pooling
- 32 Adaptive Gradient Algorithm
- 33 Root Mean Square Propagation
- 34 Adaptive Moment Estimation
- 35 Kaggle database

۵۹ ادامه در صفحه