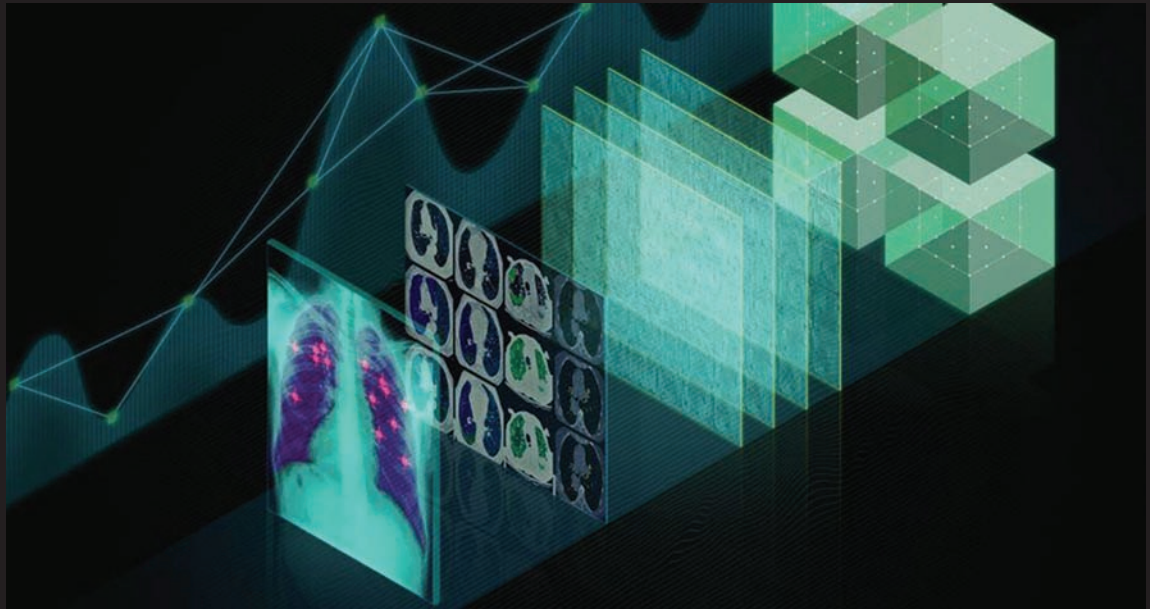


پردازش تصاویر پزشکی با استفاده از شبکه‌های عصبی کانولوشن عمیق

علی کارساز/ استادیار/گروه مهندسی برق –کنترل/ موسسه آموزش عالی خراسان/ مشهد–ایران/karsaz@khorasan.ac.ir

صبورا 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 شبکیه 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 به‌صورت 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 به‌صورت 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 سیفارنت (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 سیفارنت، 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 envolving 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 a large databas

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

[Downloaded from kiaeee.ir on 2026-02-17]

۱- مقدمه

مهم‌ترین مزیت استفاده از شبکه‌های عصبی کانولوشن[۱] (CNN) توانایی آن در استخراج ویژگی‌های تصاویر به‌صورت خودکار با استفاده از مفهوم یادگیری عمیق[۲] است[۱،۲]. به‌دلیل 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 به‌علت 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[۴،۳]. از CNNها برای تشخیص مراحل پیشرفت بیماری DR نیز استفاده می‌شود[۶،۵]. 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 کلونوسکپی[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 منافذ ریوی[۱۰] [۱۷] و تشخیص اتوماتیک آناتومی در تصاویر CT[۱۸] نام برد. کاربردهای CNNها تنها به موضوع سیستم‌های تشخیصی محدود نمی‌شود بلکه به‌تازگی 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 سیفارنت[۱۵]، الکس‌نت[۱۶] و گوگل‌نت[۱۷] معرفی و بررسی می‌شوند. درنهایت در بخش چهارم نتیجه‌گیری و جمع‌بندی ارایه می‌گردد.

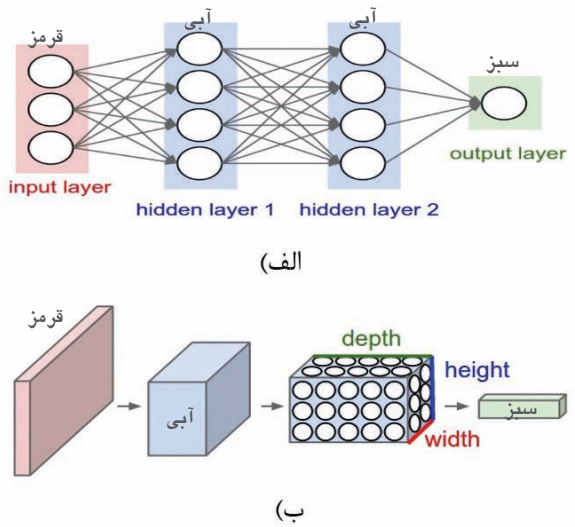
۲- شبکه‌های عصبی کانولوشن

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 CNNها 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 به‌منظور استفاده در شبکه‌های عصبی بررسی شدند. اما به‌دلیل 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 به‌طور 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 به‌طور کامل به تمامی نورون‌های لایه قبلی خود متصل می‌باشد نورون‌ها در هر لایه به‌طور 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 تنسور نیز گفته می‌شود.

در این شبکه‌ها نورون‌های یک لایه تنها به ناحیه کوچکی از لایه قبلی متصل می‌باشند و به تمامی نرون‌های لایه قبلی متصل نیستند. شکل (۱) تفاوت شبکه عصبی کانولوشن با شبکه عصبی معمولی را نشان می‌دهد. همان‌طورکه در شکل مشخص است تمامی نورون‌ها در یک TNN به هم متصل می‌باشند اما در یک CNN نورن‌ها 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].



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

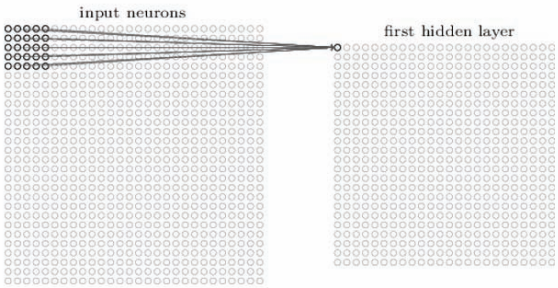
۲-۱-۱- لایه کانولوشن

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

[Downloaded from kiaeee.ir on 2026-02-17]

کانولوشـــن می‌شناسند. وظیفه ایـــن لایه استخـــراج ویژگی‌ها است. ایـــن لایه عملیات کانولوشـــن را بـــر روی داده ورودی اعمال می‌کند و خروجی‌هایـــی به نام نقشه ویژگی[۲۲] از ایـــن لایه بدست می‌آید. در نتیجـــه تمامی نورون‌هـــا در یک نقشه ویژگـــی، مجموعه‌ای از وزن‌ها و بایاس‌هـــای مشابه و مشترکی دارند کـــه باعث می‌شود، ویژگی‌های تصویـــر در موقعیت‌های مختلف قابـــل شناسایی باشند. از طرف دیگر این اشتـــراک وزن‌ها باعث کاهش تعداد پارامترهـــای مورد نیاز برای آموزش می‌گردد.

در شبکه‌هـــای کانولوشن اتصالات به‌صـــورت نواحی کوچک و محلی صـــورت می‌گیرد. به بیان دیگر هر نـــورون در نخستین لایه مخفی به ناحیه کوچکی از نورون‌های ورودی متصل است. برای مثال این ناحیه می‌توانـــد ۵×۵ باشد. ایـــن ناحیه کوچک ۲۵ پیکسلـــی ناحیه ادراک محلی[۲۳] یا کرنل کانولوشن [۲۴] نامیده می‌شود[۲۸]. برای توضیح بیشتر می‌تـــوان شکـــل(۲)، را با یک تصویـــر ورودی ۲۸×۲۸ در نظر گرفت. پنجره کرنل ۵×۵ بر روی پیکسل‌های ورودی از چپ به راست حرکت می‌کنـــد هر پنجره به نورونـــی در لایه مخفی متصل می‌شود. بنابراین همان‌طور که در شکل (۲) مشخص است لایه مخفی شامل یک شبکه ۲۴×۲۴ نورونی خواهد بود.



شکل۲ :عملیات کانولوشن در یک CNN با یک کرنل ۵×۵، راست: نورون‌های لایه پنهان، چپ: نورون‌های لایه ورودی

در شکـــل (۲) هر نورون لایه‌ی مخفی دارای یک بایاس و تعداد ۵×۵ وزن می‌باشـــد کـــه به ناحیـــه ادارکی خود متصل شـــده است. تمامی نورون‌هـــای لایه مخفی مذکـــور که دارای ابعـــاد ۲۴×۲۴ خواهد بود، دارای وزن‌ها و بایاس مشترکی می‌باشند. به بیان دیگر خروجی نورون لایه کانولوشن $y_{w,h,m}$ در طول و عرض w، h و عمق m به‌صورت رابطه (۱) قابل بیان است.

$$y_{w,h,m} = f\left(\sum_{i=(w-1)S+1}^{(w-1)S+K}\sum_{j=(h-1)S+1}^{(h-1)S+K}\sum_{k=1}^{N} W_{k,m}x_{i,j,k} + b_m\right) \quad (۱)$$

که در این رابطه f تابع فعالیت[۲۵]، b_m بایاس مشترک نورون‌ها، $W_{k,m}$ وزن‌هـــای ۵×۵ مشترک نورون‌ها و نهایتا $x_{i,j,k}$ ورودی در موقعیت j, i و k می‌باشد. بکارگیری توابع فعالیت در لایه کانولوشن، باعث افزایش خصوصیات غیرخطی در خروجی می‌شوند. توابع فعالیت انواع متفاوتی دارند. چند نمونه از معروف‌ترین توابع فعالیت، تابع سیگموئید[۲۶]، تابع تانژانـــت هایپربولیک (tanh)، تابع یکسوساز خطـــی[۲۷] (ReLU)، تابع خطـــی نمای[۲۸] (ELU) و Softmax اســـت. بنابراین تمامی نورون‌های واقـــع در لایه مخفـــی اول به‌طور دقیق ویژگـــی مشابهی را در نواحی مختلـــف تصویر شناسایی می‌نمایند. در نهایـــت خروجی لایه ورودی یا نورون‌های لایه مخفـــی به نام نقشه ویژگی شناخته می‌شوند. ابعاد مربوط به ماتریس خروجی لایه کانولوشن $W_2\times H_2\times D_2$ که از ماتریس ورودی بـــا ابعاد $W_1\times H_1\times D_1$ به‌صـــورت روابط (۲)، (۳) و (۴) به‌دست می‌آید:

$$W2 = \frac{W_۱ - F + ۲P}{S + ۱} \quad (۲)$$

$$H_۲ = \frac{H_۱ - F + ۲P}{S + ۱} \quad (۳)$$

$$D_۲ = K \quad (۴)$$

در ایـــن روابط S, P, F و K به‌ترتیب نشان‌دهنـــده اندازه کرنل، مقدار zero padding، انـــدازه‌ی گام و تعداد فیلترهـــا می‌باشد. از این روابط مشخـــص است که به‌ازای هر فیلتر به مقـــدار $F\times F\times D_1$ وزن داریم و با توجه به تعـــداد K فیلتر موجود، در مجموع، تعداد K $(F\times F\times D_1)$ وزن و K بایاس ایجـــاد می‌شود. بنابراین تعداد پارامترهایی که شبکه در یک لایه کانولوشن خود می‌باید آموزش ببیند، بسیار زیاد هستند.

۲-۱-۲-لایه پولینگ[۲۹]

در یـــک شبکـــه عصبـــی کانولوشن به‌طـــور معمول پـــس از هر لایه کانولوشن یک لایه پولینگ قرار می‌گیرد. این لایه از آن جهت اهمیت دارد که باعـــث کاهش تعداد پارامترهایی می‌شود که بایستی آموزش ببیننـــد. بنابراین با بکارگیری ایـــن لایه ضمن کاهش محاسبات مورد نیـــاز در بخش آموزش، باعث کنترل فرایادگیری[۳۰] احتمالی در شبکه نیـــز می‌گـــردد. این لایه بـــر روی هر برش از لایه‌هـــای عمق ورودی اعمال می‌شود و اندازه آن را تغییر می‌دهد. دو تابع عملکردی معروف ایـــن لایه، پولینگ بیشینه[۳۱] و میانگین[۳۲] نام دارند، که مدل نخست دارای کاربرد بیشتری در CNNها است. طریقه عمل پولینگ بیشینه بدیـــن صورت است که در هر پنجره، بزرگترین پیکسل را به خروجی می‌فرستد. این پنجره بر روی تصویر همانند تابع کانولوشن از چپ به راست و از بالا به پایین با اندازه گام‌های مشخص شده حرکت می‌کند و نتیجـــه را به خروجی می‌فرستـــد. به‌دلیل اینکه این عملیات بر روی تمامی برش‌ها اعمال می‌گـــردد، عمق خروجی همان عمق ورودی به لایه پولینـــگ می‌باشد و تغییری نمی‌کند. رابطـــه مربوط به ماتریس خروجـــی 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 می‌آید به‌صورت روابط زیر می‌باشد.

$$W_3 = \frac{W_2 - F_p}{S_p + 1} \quad (۵)$$

$$H_3 = \frac{H_2 - F_p}{S_p + 1} \quad (۶)$$

$$D_3 = D_2 \quad (۷)$$

در روابـــط بالا، Fp و Sp به‌ترتیب انـــدازه ناحیه ادراکی یا کرنل و گام لایـــه پولینگ می‌باشند. شکـــل (۳) عملکرد لایه‌ی پولینگ بیشینه را نشان می‌دهد.

۲-۱-۳-لایه‌به‌طور کامل متصل

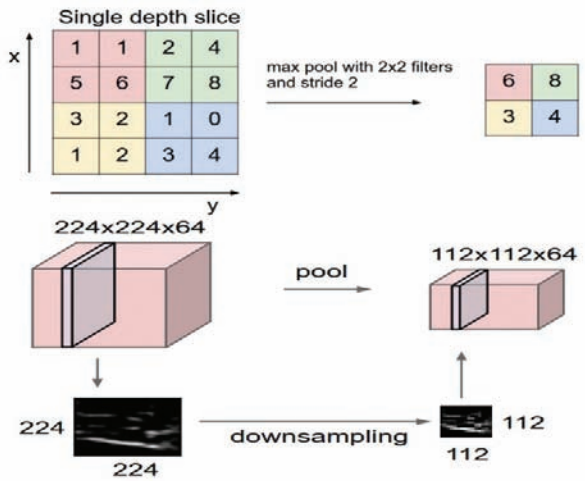
لایـــه به‌طور کامـــل متصل، لایه آخر یـــک شبکه عصبـــی کانولوشن محسوب می‌شـــود و مشابه شبکـــه عصبی معمولی اتصـــالات کاملی بـــا خروجـــی لایه قبلی ایجـــاد می‌کند. این لایـــه ورودی را دریافت و سپس خروجـــی را به‌صورت‌برداری با N مولفـــه تولید می‌کند. تعداد کلاس‌هایـــی که شبکه مورد نظـــر آن را طبقه‌بندی می‌کند برابر با N اســـت. در واقع یک شبکـــه CNN جهت تولید یک بـــردار خروجی با

[Downloaded from kiaeee.ir on 2026-02-17]

N مولفــه عددی طراحی می‌شود که هر عــدد در این بردار خروجی، درصد احتمــال تعلق به کلاس مورد نظر را نشــان می‌دهد. بالاترین درصد بیان‌گر نتیجه نهایی آموزش شبکه است[۲۶]. در الگوریتم‌های یادگیــری، بحث مهم در نظــر گرفتن رابطه‌ای بــرای بدست آوردن ارتبــاط بین کــلاس داده مورد نظــر و فرض بدست آمــده از نتیجه آمــوزش الگوریتم می‌باشد. این رابطه که با نــام تابع تلفات نیز از آن یاد می‌شــود بایستی با روش‌های مناسب، بهینه شــود. بنابراین ایده اساســی در آموزش یــک شبکه و یا الگوریتم یادگیــری، بهینه‌کردن تابــع تلفات با پیمودن گام‌های مناسب بــرای رسیدن به نتیجه بهتر اســت. روش‌های مهم بهینه‌ســازی برای یــک CNN شامل؛ گرادیان نزولــی، گرادیان نزولی batch، گرادیان نزولی تصادفی و رادیان نزولی mini-batch، الگوریتم Adagrad، الگوریتم RMSprop و Adam می‌باشد. در شبکه‌هــای عصبی کانولوشن به‌دلیــل اینکه تعداد پارامترهایی که آمــوزش می‌بینند بسیار زیاد اســت و همچنین این دسته از شبکه‌ها دارای لایه‌هــای عمیق هستند بنابرایــن در طی زمان آموزش، شبکه مستعــد فرایادگیــری می‌شود. این به این معنا اســت که شبکه برای داده‌های آموزش عملکرد خوبی از خود نشان می‌دهد اما این عملکرد بــا وارد شدن داده‌های تست دچــار مشکل می‌شود. برای حل نمودن ایــن مشکــل راه‌های متفاوت بسیــاری پیشنهاد شــده است. یکی از روش‌های معروف، روش افزایش تعداد داده‌های آموزش و روش دیگر، روش اضافــه نمودن لایه dropout مناسب در لایه به‌طور کامل متصل برای رفع این معضل است. روش افزایش داده‌ها در موضوع فرآیندهای یادگیــری، یک روش مرسوم بــوده که در آن با اعمــال تغییراتی در تصاویــر مانند چرخانــدن، بزرگ‌نمایی، جابجایی و غیــره به افزایش تعــداد تصاویر ورودی به شبکه می‌پــردازد. در روش dropout قبل از هر مرحلــه آموزش، تعــدادی از نورون‌ها به‌صورت رنــدم و با نسبت تعیین‌شــده در برنامــه مورد نظــر از فرآیند آموزش خــارج می‌شوند و تنهــا نورون‌هــای باقی‌مانده در آمــوزش شرکــت می‌کنند [۲۸].

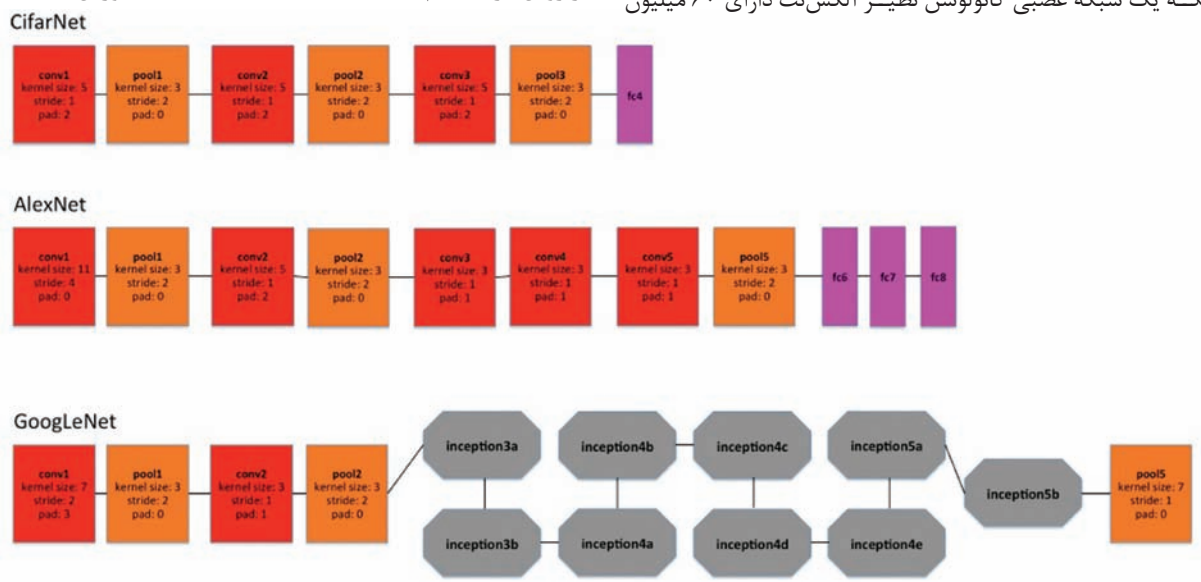
۲-۲- باز تنظیم شبکه‌های از قبل آموزش دیده

در زمینه‌های پزشکی به علت کمبود تعداد تصاویر دارای طبقه‌بندی مناسب، آمــوزش کامل یک شبکه عصبی کانولوشــن بسیار دشوار و یــا غیرممکن اســت. به‌عنوان مثال در بانــک اطلاعات چشم پزشکی کگــل تنها حــدود ۳۵ هزار تصویــر شبکیه چشم وجــود دارد حال آنکــه یک شبکه عصبی کانولوشن نظیــر آلکس‌نت دارای ۶۰ میلیون



شکل۳: عملکرد لایه پولینگ با کرنل ۲×۲ جهت کاهش پارامترهای مورد آموزش شبکه

وزن جهــت آموزش اســت. علاوه بر ایــن آموزش شبکــه از ابتدا به سخت‌افزارهــای قوی همانند پردازشگرهای قــوی گرافیکی نیاز دارد که در عمل با محدودیت سخت‌افزاری بیان شده آموزش این شبکه‌ها از ابتــدا ناممکن می‌باشــد[۲۹]. بنابراین به دو روش استخراج ویژگی و یــا استفاده از روش بازتنظیم یک شبکــه از پیش آموزش دیده در لایه‌هــای انتهایی آن می‌توان عمل نمــود. در روش استخراج ویژگی، بــه کمک یــک CNN، ویژگی‌های تصاویر پزشکــی بدست آمده و به کمک یک روش طبقه‌بندی، همانــند ماشین بردار پشتیبان می‌توان بــه دسته‌بنــدی مورد نظــر دست یافــت. این شیوه بــرای زمانی که تعــداد نمونه‌هــای بسیار کمی موجــود است دارای کاربــرد بوده در ایــن حالت بایستــی ابتدا ویژگی‌هــای تصاویر را بــه کمک لایه‌های کانولوشــن و پولینگ استخــراج نموده و در نهایت ایــن ویژگی‌ها را به‌عنــوان ورودی بــه الگوریتم‌هــای طبقه‌بندی داد، تــا طبقه‌بندی نهایی صــورت گیرد. اما زمانــی که تعداد نمونه‌هــا بیشتر است ولی به اندازه‌ای نیســت که بتوان کل پارامترهای شبکه را از ابتدا آموزش داد، و ایــن نمونه‌هــا با داده‌های آموزش شبکه اصلــی بسیار متفاوت باشــند بایستی دوباره از تنظیم یا بازتنظیم شبکه استفاده شود[۲۹]. در روش بازتنظیــم یــک CNN، ابتــدا یک شبکه آمــوزش دیده که



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

[Downloaded from kiaeee.ir on 2026-02-17]

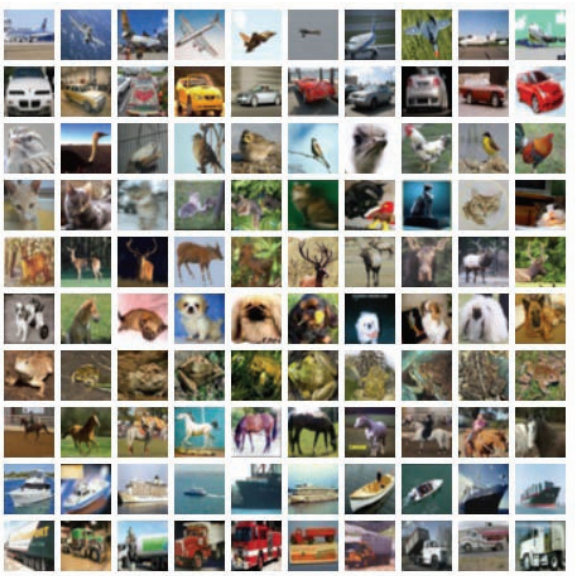
نتایـج بسیار خوبی در بانک‌های تصاویـر مطرح بدست آورده انتخاب می‌گردد، سپس لایه تماماً متصل نهایی با لایه تماماً متصل مورد نظر که قابلیت طبقه‌بندی به تعداد کلاس‌های مورد نظر را دارد، جایگزین می‌شـود. در ایـن فرآیند وزن‌های لایه‌های اولیـه شبکه اصلی بدون تغییـر وارد شبکه مورد نظر جدید می‌شونـد و لایه‌های نهایی شبکه دوباره آموزش می‌بینند تا شبکه مورد نظر قابلیت طبقه‌بندی تصاویر جدید را داشته باشد. برخی از شبکه‌های عصبی کانولوشن معروف که در چند سال اخیر معرفی شده‌اند در ادامه مورد بررسی قرار می‌گیرد.

۳- انواع شبکه‌های عصبی کانولوشن

سـه نمونـه از پرکاربردتریـن CNNهـای موجـود به‌خصـوص در حـوزه پردازش تصاویـر پزشکی در این بخش مـورد بررسی اجمالی قـرار می‌گیرند. این سه شبکـه عصبی کانولوشـن شامل شبکه‌های سیفارنـت، الکس‌نـت و گوگل‌نت بـوده که جهت بازتنظیـم آنها در تشخیص بیماری‌ها و یا بخش‌بندی‌های تصاویر پزشکی می‌توان بهره بـرد[۱۶]. شکل (۴) ساختـار کلی این سـه CNN را نشان می‌دهد.

۳-۱- شبکه CNN سیفارنت

سیفارنـت اولین بار در دانشگاه تورنتـو و در قالب تز دکتری در سال ۲۰۰۹ برای تشخیص اشیای بانک اطلاعات تصاویر سیفار۱۰ معرفی شـد[۳۰]. این بانـک اطلاعات تصاویر شامـل ۶۰۰۰۰ عکس با ابعاد ۳۲×۳۲ و در ۱۰ کـلاس مختلف بوده لذا بـه ازای هر کلاس ۶۰۰۰ عکس متفاوت کوچک وجـود دارد. شکل (۵) نمونه‌ای از تصاویر این بانک اطلاعاتـی را نشان می‌دهد. سیفارنت دارای سه لایه کانولوشن، سه لایه پولینگ و یک لایه تماما متصل می‌باشد.



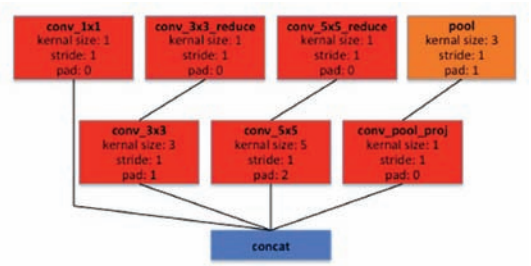
شکل ۵: چند نمونه از تصاویر بانک اطلاعاتی سیفار۱۰، شامل ۶۰۰۰۰ عکس در ۱۰ کلاس متفاوت با ابعاد ۳۲×۳۲

۳-۲- شبکه CNN الکس‌نت

این شبکه در ۲۶امین کنفرانس سیستم‌های پردازش اطلاعات عصبی در سـال ۲۰۱۲ معرفـی گردیـد[۲۶]. این شبکه برتـری خود را بر شبکه‌های عصبی غیر عمیق برای تشخیص تصاویر در مسابقه معروف تشخیـص تصاویر ابعاد بزرگ ImageNet2 به اثبات رساند[۳۱]. بانک اطلاعات ImageNet شامل ۱/۲ میلیون تصویر با ابعاد ۲۵۶×۲۵۶ بوده کـه در ۱۰۰۰ کلاس مختلف گردآوری شده‌اند. شکل (۶) چند نمونه از تصاویر این بانک اطلاعات تصاویر را نشان می‌دهد. الکس‌نت شامل ۵ لایه کانولوشن، سه لایه پولینگ و دو لایه تماما متصل بوده و حدود ۶۰ میلیون پارامتر آزاد جهت آموزش دارد.

۳-۳- شبکه CNN گوگل‌نت و مدل Inception-v[3]

شبکـه عصبـی کانولوشـن گوگل‌نـت به‌عنـوان عمیق‌تریـن شبکه CNN در بیـن مدل‌هـای رایـج بـوده مهمتریـن وجـه تمایـز این شبکـه نسبـت بـه مدل‌هـای قبلـی ماننـد الکس‌نت یـا سیفارنت استفـاده از مـاژول Inception در ساختـار خـود اسـت[۱]. بـه شکل کلـی یـک شبکـه کانولوشـن گوگل‌نـت دارای دو لایه کانولوشـن، دو لایـه پولینـگ و ۹ لایـه Inception بـوده هـر لایه Inception خـود دارای ۶ لایـه کانولوشـن و یـک لایـه پولینـگ اسـت. شکـل (۷) ساختار یـک گوگل‌نـت را نشان می‌دهـد. مدل Inception-v[3] به‌عنـوان یک مدل پیشرفته از شبکه عصبی کانولوشن گوگل‌نـت، اولیـن بار توسـط Szegedy و همکـاران در سال ۲۰۱۵ معرفـی گردید [۲]. در این مدل، کانولوشن‌های ۱×۱، ۳×۳ و ۵×۵ در یـک ماژول یکسان و بر روی یـک ورودی اعمال می‌شوند و خروجی تمامـی این فرآیندهـا با یکدیگر الحاق شده تـا به‌عنوان ورودی لایه بعـدی در نظر گرفته شوند. شکل (۸) یک ماژول Inception از شبکه کانولوشـن گوگل‌نت را نشـان می‌دهد. به دلی-ل الحـاق مذکور نام دیگر مـاژول Inception، استخراج‌کننده‌ی ویژگی چندلایه می‌باشد.



شکل ۸: یک مـاژول Inception-a3 از شبکه عصبی کانولوشن گوگل‌نت

۴- نتیجه‌گیری

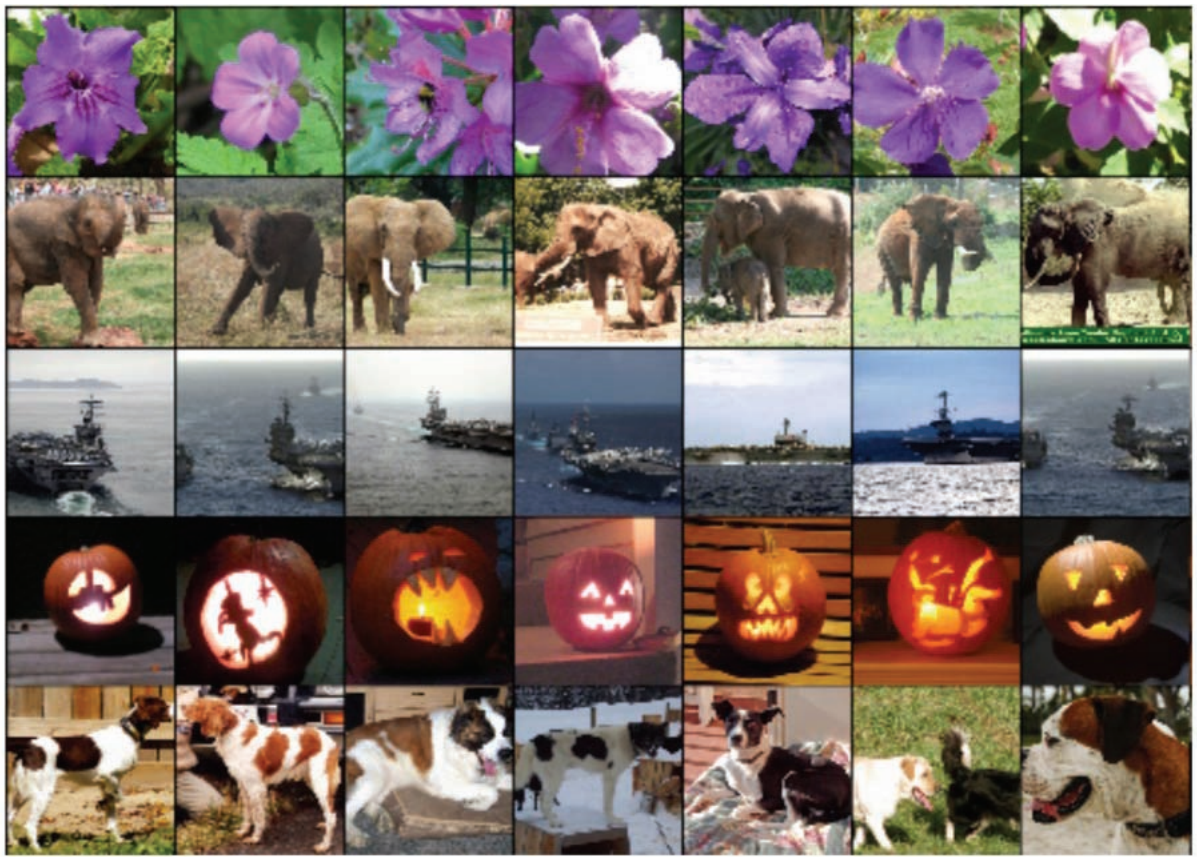
شبکه‌هـای عصبی مصنوعـی کانولوشن در طول کمتـر از یک دهه گذشته با توسعه قدرت پردازش کامپیوترها، در حوزه یادگیری عمیق سهم به‌سزایی را ایفا نموده به‌گونه‌ای که در کاربردهای مهم و حیاتی ماننـد پـردازش تصاویر پزشکـی و سیستم‌های اتومبیـل خودران با موفقیت‌های روزافزونی همراه بوده‌اند.

بکارگیـری شبکه‌هـای کانولوشـن از قبیـل سیفارنـت، گوگل‌نت و الکس‌نـت به‌علـت استفاده از لایه‌هـای عمیق بیشتـر در مقایسه با شبکه‌هـای عصبی متـداول، می‌تواند در پـردازش تصاویر پزشکی و مقوله غربالگری بیماری‌ها تحول چشم‌گیری را ایجاد نماید.

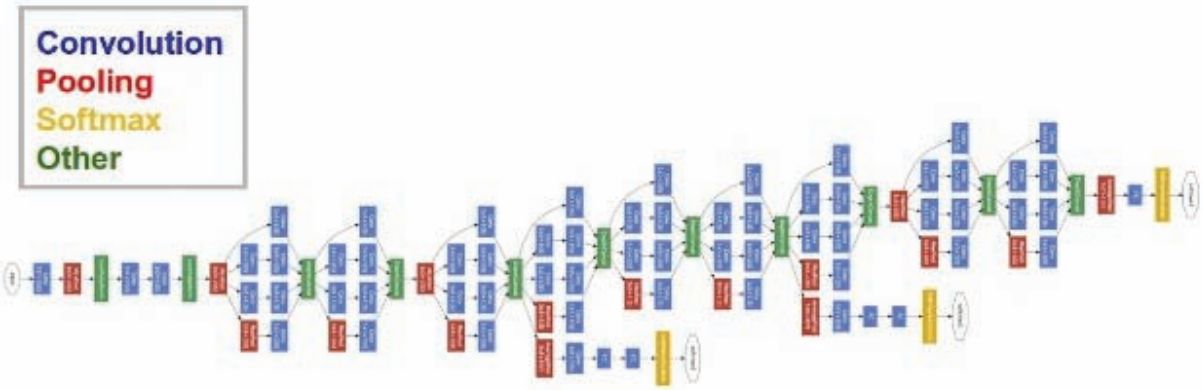
پی‌نوشت‌ها

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

[Downloaded from kiaeee.ir on 2026-02-17]



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 به‌عنوان عمیق‌ترین CNN موجود

15 CifarNet
16 AlexNet
17 GoogleNet
18 Traditional neural network
19 Fukushima
20 pooling layer
21Fully 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
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. Wanga, G. Caoa, B. Weib 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. Broadbentc, P. Simon P. Hardingac and Y. Zhengac, "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

ادامه در صفحه ۵۹

[Downloaded from kiaeee.ir on 2026-02-17]