

۲. معماری یکنواخت: معماری‌های VGG از یک ساختار ثابت پیروی می‌کنند که درک و پیاده‌سازی آن‌ها را آسان‌تر می‌کند.

۳. استخراج موثر ویژگی: با استفاده از فیلترهای کوچک 3×3 که روی هم چیده شده‌اند، VGG می‌تواند به طور موثر ویژگی‌های سلسله مراتبی را از تصاویر استخراج کند و تشخیص قوی تصویر را امکان پذیر می‌کند.

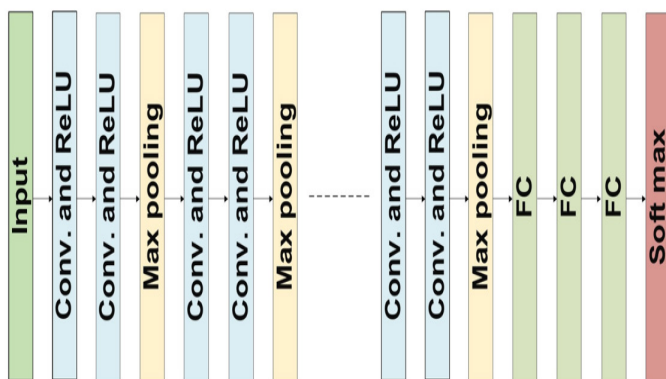
۴. سادگی: معماری‌های VGG علی‌رغم عمقشان، طراحی ساده‌ای دارند، که اصلاح و تطبیق آن‌ها را برای وظایف یا مجموعه داده‌های خاص آسان می‌کند.

معایب مدل :

۱. هزینه محاسباتی بالا: مدل‌های VGG دارای تعداد زیادی پارامتر به خصوص VGG19 هستند که می‌تواند منجر به نیازهای محاسباتی بالا در طول آموزش و استنتاج شود.

۲. حافظه فشرده: تعداد زیاد پارامترها در مدل‌های VGG همچنین می‌تواند آن‌ها را به حافظه فشرده تبدیل کند و به منابع قابل توجهی برای آموزش و استقرار نیاز دارد.

۳. اطلاعات متنی محدود: به دلیل استفاده از فیلترهای کوچک 3×3 در سراسر شبکه، VGG ممکن است در مقایسه با معماری‌هایی با میدان‌های دریافتی بزرگ‌تر، برای گرفتن اطلاعات زمینه‌ای دوربرد در تصاویر دچار مشکل شود.



معماری مدل VGG

مدل ResNet (Residual Network)

ResNet یک معماری شبکه عصبی کانولوشن عمیق است که برای رفع چالش آموزش شبکه‌های بسیار عمیق معرفی شده است. نوآوری کلیدی آن در استفاده از بلوک‌های باقیمانده است که اتصالات میانبر یا اتصالات را رد می‌کنند. این اتصالات اجازه می‌دهد ورودی مستقیماً به لایه‌های عمیق‌تر منتقل شود و آموزش شبکه‌های عمیق را با امکان یادگیری توابع باقیمانده تسهیل می‌کند. معماری ResNet شامل چندین مرحله است که هر کدام شامل چندین بلوک باقیمانده است. شبکه معمولاً با لایه‌های کانولوشن استاندارد شروع می‌شود و سپس مراحل بلوک‌های باقیمانده دنبال می‌شود. نمونه برداری پایین با استفاده از لایه‌های کانولوشن با گام ۲ به دست می‌آید و ابعاد فضایی را کاهش می‌دهد و تعداد کانال‌ها را افزایش می‌دهد. این معماری انواع مختلفی دارد، مانند ResNet-5، ResNet-50، ResNet-101، ResNet-152، با عمق‌ها و پیچیدگی‌های متفاوت. مدل‌های ResNet عملکرد فوق‌العاده‌ای را در

شبکه عصبی کانولوشنال در یک نگاه

شبکه‌های عصبی کانولوشنال (CNN) دسته‌ای از مدل‌های یادگیری عمیق هستند که به طور خاص برای پردازش داده‌های شبکه ساختاریافته، به‌ویژه تصاویر طراحی شده‌اند. CNN‌ها که از لایه‌هایی تشکیل شده است که نمایش‌های سلسله‌مراتبی را از طریق فیلترهای کانولوشنال استخراج می‌کند، با یادگیری خودکار ویژگی‌های مرتبط از داده‌های پیکسل خام، وظایف بینایی رایانه را متحول کرده است. با استفاده از تکنیک‌هایی مانند ادغام لایه‌ها برای نمونه برداری فضایی و توابع فعال‌سازی برای معرفی غیرخطی، CNN‌ها می‌توانند به طور موثر روابط پیچیده را در تصاویر مدل کنند. معماری‌های محبوب CNN مانند ResNet، VGG، Inception و ResNet عملکرد قابل توجهی را در برنامه‌های



امیر مختار راد
دانشجو مهندسی کامپیوتر
دانشکده فابرای دانشگاه تهران
amir.r.rad@ut.ac.ir

مختلف، از طبقه بندی تصویر و تشخیص اشیاء گرفته تا تقسیم بندی معنایی، نشان داده‌اند. از طریق آموزش تکراری با استفاده از مجموعه داده‌های برچسب‌گذاری شده، و تکنیک‌های بهینه‌سازی، CNN‌ها پارامترهای خود را برای به حداقل رساندن خطاهای پیش‌بینی تطبیق می‌دهند و آنها را به ابزاری ارزشمند برای کارهایی تبدیل می‌کنند که نیاز به درک و تفسیر بصری دارند. در ادامه به معرفی و نقد چند نمونه از مدل‌های CNN می‌پردازیم :

مدل VGG (Visual Geometry Group)

پس از اینکه CNN‌ها در تشخیص تصویر موثر بودند، Simonyan و Zisserman یک اصل طراحی کارآمد به نام (Visual Geometry Group) را پیشنهاد کردند. این مدل نوآورانه دارای دو نوع محبوب است: VGG16 و VGG19 که به ترتیب با عمق ۱۶ و ۱۹ لایه متمایز می‌شوند. برخلاف فیلترهای بزرگ‌تر مورد استفاده در رویکردهای قبلی، VGG از پشته‌ای از فیلترهای 3×3 استفاده کرد، که به صورت تجربی نشان داد که می‌تواند نتایج مشابهی را در حالی که تعداد پارامترها و پیچیدگی محاسباتی را کاهش می‌دهند، به دست آورند. علاوه بر این، VGG پیچیدگی‌های 1×1 را برای تنظیم پیچیدگی شبکه و حداکثر لایه‌های ادغام به دنبال لایه‌های کانولوشن برای حفظ وضوح فضایی درج کرد. VGG علی‌رغم برنده نشدن در رقابت ILSVRC 2014، به دلیل افزایش عمق، توپولوژی یکنواخت و سادگی شهرت پیدا کرد. با این حال، اشکال اصلی آن هزینه محاسباتی بالای آن بود که ناشی از استفاده از حدود ۱۴۰ میلیون پارامتر است.

مزایای مدل :

۱. افزایش عمق: VGG16 و VGG19 در مقایسه با مدل‌های قبلی دارای لایه‌های بیشتری هستند که به آنها امکان می‌دهد ویژگی‌های پیچیده‌تری را از تصاویر ورودی ثبت کنند.

وظایف تشخیص تصویر از خود نشان داده‌اند و باعث شده است که آنها به طور گسترده در برنامه‌های بینایی رایانه مورد استفاده قرار گیرند. مزایای مدل :

۱. آموزش بهبود یافته: استفاده ResNet از اتصالات پرش امکان آموزش آسان تر شبکه‌های بسیار عمیق را با کاهش مشکل گرادیان ناپدید می‌کند که منجر به همگرایی بهبود یافته و عملکرد بهتر می‌شود.

۲. افزایش عمق: معماری‌های ResNet می‌توانند به طور قابل توجهی عمیق تر از مدل‌های قبلی باشند بدون اینکه از کاهش عملکرد رنج ببرند، و اجازه می‌دهد تا نمایش‌های پیچیده تری یاد بگیرند.

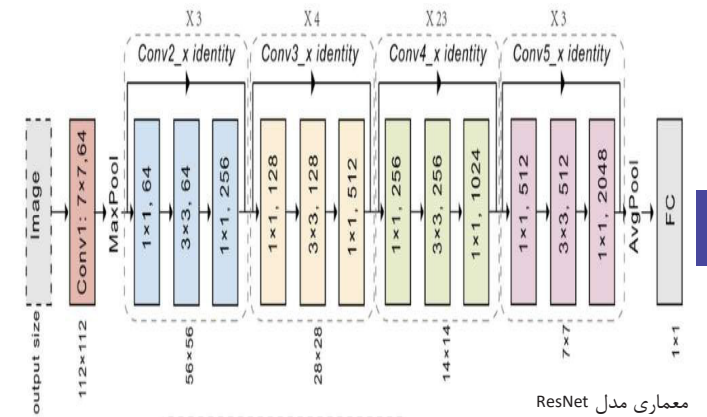
۳. عملکرد پیشرفته: مدل‌های ResNet در معیارهای مختلف تشخیص تصویر، از جمله چالش تشخیص تصویر در مقیاس بزرگ (ImageNet ILSVRC) به نتایج پیشرفته‌ای دست یافته‌اند که کارایی خود را در عمل نشان می‌دهد.

معایب مدل :

۱. پیچیدگی محاسباتی: معماری‌های عمیق تر ResNet به منابع محاسباتی بیشتری برای آموزش و استنتاج در مقایسه با مدل‌های کم‌عمق تر نیاز دارند، که به طور بالقوه کاربرد آنها را در محیط‌های با محدودیت منابع محدود می‌کند.

۲. تطبیق بیش از حد: در حالی که اتصالات پرش ResNet به کاهش مشکل گرادیان ناپدید شدن کمک می‌کند، ممکن است خطر بیش از حد برازش را نیز ایجاد کند، به خصوص در معماری‌های بسیار عمیق با تعداد زیادی پارامتر.

۳. تفسیرپذیری: افزایش عمق و پیچیدگی مدل‌های ResNet ممکن است آنها را در مقایسه با معماری‌های ساده‌تر تفسیرپذیرتر کند و چالش‌هایی را برای درک بازنمایی‌های آموخته‌شده و فرآیندهای تصمیم‌گیری ایجاد کند.

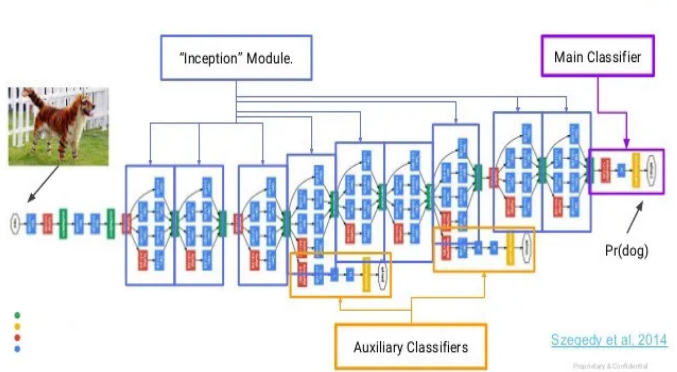


مدل (Inception) GoogleNet :

در رقابت ILSVRC ۲۰۱۴، GoogleNet (همچنین به عنوان Inception-V1 شناخته می‌شود) به عنوان برنده ظاهر شد و بر دستیابی به دقت بالا با کاهش هزینه محاسباتی تمرکز کرد. این مفهوم بلوک آغازین را معرفی کرد که تبدیل‌های کانولوشنال چند مقیاسی را از طریق توابع ادغام، تبدیل و تقسیم برای استخراج ویژگی ترکیب می‌کند. این معماری دارای فیلترهایی با اندازه‌های مختلف (۱x۱، ۳x۳، ۵x۵) برای ضبط کانال و

اطلاعات مکانی در وضوح‌های مختلف است. GoogleNet لایه کانولوشنال رایج را با بلوک‌های کوچک الهام گرفته از معماری شبکه در شبکه (NIN) جایگزین کرد و از اتصالات پراکنده برای رسیدگی به اطلاعات اضافی و کاهش محاسبات استفاده کرد. GoogleNet با استفاده از یک لایه گلوگاه با پیش‌پیش‌های ۱x۱ و یک لایه میانگین جهانی (GAP) به جای لایه‌های کاملاً متصل، تعداد پارامترها را به طور قابل توجهی به ۵ میلیون کاهش داد. همچنین تکنیک‌های منظم‌سازی مانند بهینه‌ساز RmsProp و نرمال‌سازی دسته‌ای را همراه با یادگیرنده‌های کمکی برای تسریع در همگرایی معرفی کرد. با این حال، توپولوژی ناهمگن GoogleNet چالش‌هایی را برای انطباق بین ماژول‌ها ایجاد می‌کند و پارازیت نمایش آن گاهی اوقات منجر به از دست دادن اطلاعات در لایه‌های بعدی می‌شود. معماری Inception، که با Inception-V1 (GoogleNet) شروع شد، از طریق تکرارهایی مانند Inception-V2، Inception-V3، و Inception-V4 ResNet تکامل یافت. هدف این نسخه‌ها افزایش کارایی محاسباتی و عملکرد تعمیم با ترکیب ویژگی‌هایی مانند کانولوشن‌های فاکتوریزه، نرمال‌سازی دسته‌ای و اتصالات باقی مانده بود. هر تکرار نشان‌دهنده تلاش‌های مداوم برای بهینه‌سازی طرح‌های شبکه عصبی کانولوشن برای بهبود عملکرد در وظایف مختلف بینایی رایانه است.

GoogLeNet (aka "Inception") Architecture



مزایای مدل :

۱. دقت بالا: مدل‌های Inception، به‌ویژه نسخه‌های جدیدتر مانند Inception-V3 و فراتر از آن، در کارهای مختلف بینایی کامپیوتر، از جمله طبقه‌بندی تصویر و تشخیص اشیاء، به دقت بالایی دست می‌یابند.

۲. استفاده کارآمد از پارامترها: با ترکیب تکنیک‌هایی مانند پیش‌پیش‌های فاکتوریزه و منظم‌سازی تهاجمی، مدل‌های Inception به طور موثر از پارامترها استفاده می‌کنند و عملکرد بهتری را با پارامترهای کمتر از مقایسه با برخی معماری‌های دیگر ممکن می‌سازند.

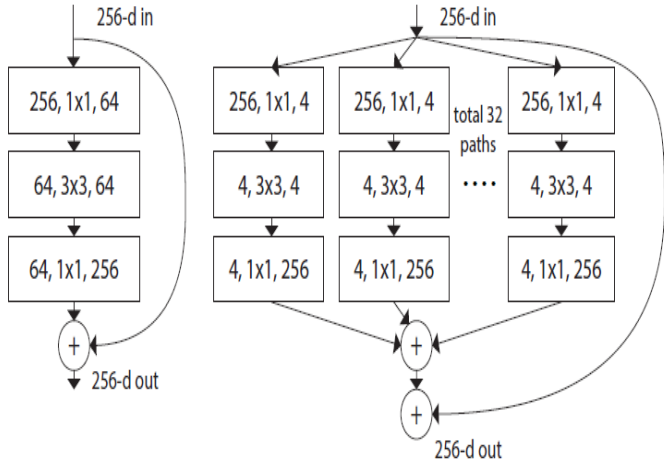
۳. انعطاف‌پذیری: مدل‌های Inception همه‌کاره هستند و می‌توانند برای وظایف و مجموعه داده‌های مختلف با تنظیم دقیق یا انتقال یادگیری تطبیق داده شوند و آنها را برای طیف وسیعی از کاربردها مناسب می‌سازد.

معایب مدل :

۱. پیچیدگی محاسباتی: با وجود کارایی، مدل‌های Inception هنوز هم می‌توانند محاسباتی فشرده باشند، مخصوصاً نسخه‌های بزرگ‌تر، که به

و استنتاج، به ویژه برای مدل ها و مجموعه داده های بزرگتر نیاز داشته باشد.

۳. پایداری تمرین: آموزش مدل های ResNeXt ممکن است به تنظیم دقیق فرآیندها و تکنیک های آموزشی برای اطمینان از پایداری و جلوگیری از مسائلی مانند بیش از حد نیاز داشته باشد.



معماری مدل ResNeXt

مقایسه ای بین چهار مدل معرفی شده در کاربرد آن ها :

در استفاده روزانه و برنامه های مختلف، مدل های VGG، ResNet، Inception و ResNeXt مزایای مشخصی را ارائه می دهند. استحکام VGG در طبقه بندی تصاویر آن را برای کارهایی مانند بازیابی تصویر مبتنی بر محتوا، تجزیه و تحلیل تصویر رسانه های اجتماعی و سازماندهی عکس مناسب می کند. عملکرد استثنایی ResNet در معماری های بسیار عمیق، آن را قادر می سازد تا وظایف پیچیده ای مانند تشخیص دقیق تصویر، تشخیص اشیاء سه بعدی و تجزیه و تحلیل ویدئو را انجام دهد و آن را در تجزیه و تحلیل تصاویر ماهواره ای، اتوماسیون صنعتی و مراقبت های بهداشتی ارزشمند می سازد. کارایی محاسباتی Inception برای کاربردهای بلادرنگ مانند تشخیص تصویر موبایل، فیلترهای واقعیت افزوده و موتورهای جستجوی تصویر ایده آل است و راحتی را در زمینه های مختلف افزایش می دهد. کاردینالیته بهبود یافته و بلوک های چند تبدیلی ResNeXt دقت و کارایی بهبود یافته ای را ارائه می کند و آن را برای برنامه هایی که نیاز به دقت سطح بالا با کاهش هزینه محاسباتی دارند، مانند تصویربرداری پزشکی و وسایل نقلیه خودران، مناسب می سازد. به طور کلی، هر مدل نقاط قوت منحصر به فردی را ارائه می دهد که طیف گسترده ای از موارد استفاده روزانه و برنامه های کاربردی را در حوزه های مختلف ارائه می دهد.

منابع :

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., Farhan, L (2021). A Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, and Future Directions. Journal of Big Data, Volume (53), (5).

Convolutional Neural Network: Dhillion, A., Verma, G. K (2019). A Review of Models, Methodologies, and Applications to Object Detection. Progress in Artificial Intelligence, Volume (57), (5).

منابع محاسباتی قابل توجهی برای آموزش و استنتاج نیاز دارند.

۲. پیچیدگی تمرین: تمرین مدل های Inception ممکن است به تنظیم دقیق فرآیندها و تکنیک های منظم سازی برای جلوگیری از برازش بیش از حد و اطمینان از هم گرایی نیاز داشته باشد که به فرآیند آموزش پیچیدگی می افزاید.

۳. نیازهای حافظه: مدل های Inception، به ویژه نسخه های بزرگ تر، ممکن است در حین آموزش و استنباط نیاز به حافظه بالایی داشته باشند، که نیاز به منابع حافظه کافی برای اجرای کارآمد دارند.

مدل ResNeXt :

ResNeXt نشان دهنده انحراف از هر دو معماری VGG و Inception است و رویکردی منحصر به فرد برای مدل های یادگیری عمیق ارائه می دهد. برخلاف VGG که بر عمق با توپولوژی همگن تاکید دارد و Inception که بر تبدیل های کانولوشنال در مقیاس چندگانه تمرکز دارد، ResNeXt مفهوم کاردینالیته را برای افزایش استخراج ویژگی معرفی می کند. این امکان استفاده موثرتر از منابع شبکه و بهبود توانایی یادگیری را در مقایسه با CNN های سنتی فراهم می کند. ResNeXt با ترکیب ویژگی های معماری های VGG، ResNet، و Inception، با استفاده از یک توپولوژی همگن عمیق شبیه به VGG و در عین حال ترکیب ساختار اصلی GoLeNet، به این امر دست می یابد. استفاده از آن شامل استفاده از تبدیل های چندگانه در داخل بلوک های تقسیم، تبدیل و ادغام است، در نتیجه عملکرد را از طریق افزایش کاردینالیته بهینه می کند. این رویکرد تعادلی بین عمق و کارایی محاسباتی ارائه می کند که با استفاده استراتژیک از فیلترهای 1x1 و آموزش بهینه شده از طریق پرش از اتصالات تنظیم می شود. به این ترتیب، ResNeXt یک گزینه همه کاره و موثر برای کارهای مختلف بینایی کامپیوتری است که جایگزینی منحصر به فرد برای VGG متمرکز بر عمق و معماری های Inception چند مقیاسی است.

مزایای مدل :

۱. استخراج ویژگی های پیشرفته: معرفی مفهوم کاردینالیته به ResNeXt اجازه می دهد تا به طور موثر از منابع شبکه استفاده کند و نمایش ویژگی های غنی تر را از داده های ورودی ضبط کند.
۲. عمق و کارایی متوازن: ResNeXt تعادلی بین عمق و کارایی محاسباتی ایجاد می کند و جایگزینی همه کاره برای معماری هایی مانند VGG و Inception ارائه می دهد.

۳. عملکرد بهینه شده: استفاده از بلوک های چند تبدیلی و بهینه سازی های استراتژیک مانند فیلترهای 1x1 و پرش از اتصالات منجر به بهبود عملکرد در وظایف مختلف بینایی رایانه می شود.

معایب مدل :

۱. پیچیدگی: ResNeXt ممکن است در مقایسه با معماری های ساده تر مانند VGG برای درک و پیاده سازی پیچیده تر باشد و به طور بالقوه به تخصص بیشتری در یادگیری عمیق نیاز دارد.
۲. منابع محاسباتی: در حالی که ResNeXt برای کارایی طراحی شده است، ممکن است همچنان به منابع محاسباتی قابل توجهی برای آموزش