

به نام خدا



تمرین ششم یادگیری عمیق

استاد درس : دکتر مرضیه داوودآبادی

سارا سادات یونسی _ ۹۸۵۳۳۰۵۳

فهرست

سوال ۱.....	صفحه ۳
سوال ۲.....	صفحه ۶
سوال ۳.....	صفحه ۸
سوال ۴.....	صفحه ۱۳
سوال ۵.....	صفحه ۱۸

سوال ۱

(الف)

شبکه کانولوشنال (CNN) نوعی شبکه عصبی مصنوعی است که می‌تواند تصاویر را با اعمال فیلترها یا هسته‌های متعدد برای استخراج ویژگی‌ها پردازش کند. مکانیسم توجه روشی است که می‌تواند بر روی قسمت‌های خاصی از یک تصویر یا یک صحنه تمرکز کند و در عین حال قسمت‌های نامربوط را نادیده بگیرد.

CNN ها و مکانیسم‌های توجه هر دو می‌توانند حضور گریه‌ها را در تصویر تشخیص دهند، حتی اگر تصویر دارای تضاد نشانه بافتی باشد، مانند شکل گریه با پوست فیل. با این حال، آنها ممکن است از استراتژی‌های متفاوتی استفاده کنند و سطح دقت متفاوتی داشته باشند.

CNN ها می‌توانند با استفاده از سلسله مراتبی از ویژگی‌ها و یک لایه طبقه‌بندی، شناخت گریه‌ها را بیاموزند. آنها می‌توانند ویژگی‌های سطح پایین مانند لبه‌ها و گوشه‌ها را استخراج کنند و آنها را با ویژگی‌های سطح بالا مانند چشم، گوش و سبیل ترکیب کنند. با این حال، CNN ها ممکن است نتوانند ساختار جهانی و زمینه تصویر و همچنین مکانیسم‌های توجه را به تصویر بکشند. آنها همچنین ممکن است با تضاد نشانه بافت-شکل اشتباه گرفته شوند، به خصوص اگر بافت غالب یا مشابه پس زمینه باشد. به عنوان مثال، یک CNN ممکن است یک گریه با پوست فیل را بسته به کیفیت تصویر و داده‌های آموزشی به اشتباه طبقه‌بندی کند، به عنوان فیل یا سنگ.

مکانیسم‌های توجه می‌توانند با استفاده از دانش و انتظارات قبلی در مورد ساختار و ظاهر گریه‌ها، شناخت گریه‌ها را بیاموزند. آنها می‌توانند روی قسمت‌های مرتبط تصویر، مانند شکل و ویژگی‌های صورت، تمرکز کنند و بخش‌های نامربوط مانند بافت و پس‌زمینه را نادیده بگیرند. آنها همچنین می‌توانند از زمینه و کنتراست برای تشخیص گریه‌ها در یک تصویر مانند شکل و رنگ پوست، مو و پس‌زمینه استفاده کنند. با این حال، مکانیسم‌های توجه ممکن است نتوانند جزئیات و تغییرات ریز دانه گریه‌ها و همچنین CNN ها را استخراج کنند. آنها همچنین ممکن است تحت تأثیر برجسته بودن و تازگی تضاد نشانه بافت-شکل قرار گیرند، به خصوص اگر شکل غالب یا غیر معمول باشد. به عنوان مثال، یک مکانیسم توجه ممکن است بسته به وظیفه و هدف، توجه بیشتری به گریه‌ای با پوست فیل نسبت به گریه معمولی داشته باشد.

به طور خلاصه، CNN ها و مکانیسم‌های توجه هر دو می‌توانند گریه‌ها را در یک تصویر تشخیص دهند، حتی اگر تصویر دارای تضاد نشانه بافت شکل باشد، اما ممکن است نقاط قوت و ضعف متفاوتی داشته باشند. CNN ها می‌توانند از ویژگی‌های آموخته

شده و وزن های طبقه بندی برای تشخیص گربه ها استفاده کنند، در حالی که مکانیسم های توجه می توانند از دانش، انتظارات، زمینه و تضاد قبلی برای تشخیص گربه ها استفاده کنند.

هر دوتصویر عادی را می توانند به درستی تشخیص دهند اما درباره ی تصویر خاکستری با خطوط فیل احتمال اشتباه کردن CNN بالاست و مکانیزم توجه می تواند بهتر عمل کند با توجه به اینکه به روابط پیکسل ها با یکدیگر در همه جای تصویر توجه می کند. CNN ها اگر درصد تکستچر بالا باشد توانایی تشخیص ندارند این آزمایش در RESNET که دارای کانولوشنال های موازی زیادی هم می باشد ۶۰ درصد فیل افریقایی تشخیص داده است اما مکانیزم توجه می تونه به بخش های خاصی از تصویر توجه بکند و پس زمینه را در نظر نگیرد که دارای اطلاعات غیر مرتبط است.

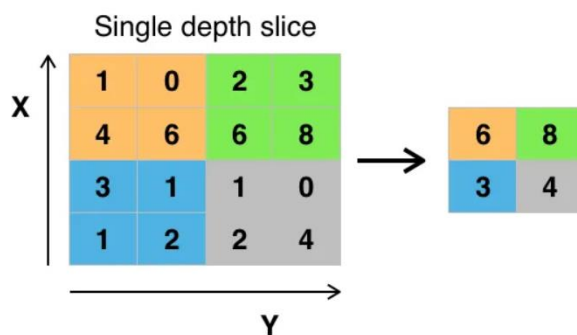
<https://medium.com/@robertgeirhos/why-deep-learning-works-differently-than-we-thought-ec28823bdbc>

ب

شبکه کانولوشنی هر دو شکل را انسان در نظر می گیرد.

در اصل، CNN ها با تشخیص وجود ویژگی ها در یک تصویر، و استفاده از دانش این ویژگی ها برای پیش بینی وجود یک شی کار می کنند. با این حال، سی ان ان ها فقط وجود ویژگی ها را تشخیص می دهند – بنابراین حتی تصویر سمت راست که شامل چشم، بینی و گوش نامناسب است، یک چهره در نظر گرفته می شود، زیرا شامل تمام ویژگی های لازم است!

مشکل به نحوه انتقال اطلاعات از لایه ای به لایه دیگر توسط CNN ها خلاصه می شود: حداکثر ادغام. Max Pooling شبکه ای از پیکسل ها را بررسی می کند و حداکثر فعال سازی را در آن ناحیه انجام می دهد. این روش اساساً تشخیص می دهد که آیا یک ویژگی در هر منطقه ای از تصویر وجود دارد یا خیر، اما اطلاعات مکانی مربوط به ویژگی را از دست می دهد.



Max Pooling preserves existence (maximum activation) in each region, but loses spatial information. Source.

شبکه‌های مبتنی بر توجه سعی می‌کنند محدودیت‌های CNN را با گرفتن روابط جزئی از کل برطرف کنند. به عنوان مثال، در یک صورت، هر یک از دو چشم بخشی از یک پیشانی است. برای درک این روابط جزئی، مکانیزم نیاز به دانستن بیشتر از وجود یک ویژگی دارند - باید بدانند ویژگی کجاست، چگونه جهت‌گیری می‌کند و سایر اطلاعات مشابه. به روش‌هایی موفق می‌شوند که CNN ها موفق نشده‌اند، زیرا آنها با موفقیت این اطلاعات را دریافت می‌کنند و اطلاعات را از قسمت‌ها به کل خود هدایت می‌کنند.

یک راه حل موجود برای این مشکل شبکه کیسولی می باشد

ایده این است که برای اضافه کردن ساختارهایی به نام کپسول به شبکه‌های عصبی کانولوشن (CNN) و به استفاده مجدد خروجی از چند تا از این کپسول به فرم با ثبات‌تر (با توجه به آشفتگی) قابل ارائه برای کپسول‌های مرتبه بالاتر. خروجی یک وکتور متشکل از احتمال مشاهده و ژست (موقعیت، اندازه، جهت) که برای مشاهده. این وکتور شبیه به آنچه که برای مثال در هنگام انجام طبقه‌بندی یا محله سازی در CNN ها انجام می‌شود است.

اما با مکانیزم توجه می توان تصویر درست از غلط انسان را شناسایی کرد و فقط چهره ی سمت چپ به عنوان چهره ی انسان شناسایی شود.

مکانیزم توجه می تواند در موارد مختلف پردازش تصویر مانند تشخیص چهره، شناسایی اشیاء، تولید شرح برای تصاویر و ترجمه تصویر به متن استفاده شود. در این موارد، مکانیزم توجه به شبکه کمک می کند تا روی اشیاء یا ویژگی های خاصی در تصویر خاص کند و از این طریق ویژگی ها و اطلاعات تصویر را می کند. مکانیزم توجه می تواند به صورت داخلی (self-attention) یا خارجی (cross-attention) باشد.

attention) باشد. در حالت داخلی، بر روی شبکه های مختلف یک تصویر خارجی می کند و در حالت، شبکه بر روی بخش های مختلف دو تصویر متفاوت متفاوت می کند

استفاده از مکانیزم توجه در داده های تصویری، یک روش است که برای بهبود عملکرد شبکه های عصبی کانولوشنی (CNN) در پردازش تصاویر استفاده می شود. این روش با اجازه دادن به شبکه برای روی روی بخش های مهم و مرتبط با تصویر در حین تولید پیش بینی استفاده می شود. این کار با وزن دادن به فیلتر ها یا کرنل های شبکه انجام می شود تا شبکه برخی از ویژگی های تصویر را بر سایرین اولویت دهد. در نتیجه، شبکه می تواند پیش بینی های دقیق تری را تنها با در نظر گرفتن مهم ترین تغییرات تصویر تولید کند.

مکانیزم توجه می تواند در موارد مختلف پردازش تصویر مانند تشخیص چهره، شناسایی اشیاء، تولید شرح برای تصاویر و ترجمه تصویر به متن استفاده شود. در این موارد، مکانیزم توجه به شبکه کمک می کند تا روی اشیاء یا ویژگی های خاصی در تصویر خاص کند و از این طریق ویژگی ها و اطلاعات تصویر را می کند. مکانیزم توجه می تواند به صورت داخلی (self-attention) یا خارجی (cross-attention) باشد. در حالت داخلی، بر روی شبکه های مختلف یک تصویر خارجی می کند و در حالت، شبکه بر روی بخش های مختلف دو تصویر متفاوت متفاوت می کند

<https://towardsdatascience.com/demystifying-matrix-capsules-with-em-routing-part-1-overview-2126133a8457>

سوال ۲

(الف)

		Actual Values	
		Yes	No
Predicted Values	Yes	True Positive	False Positive
	No	False Negative	True Negative

ساختار ماتریس در هم ریختگی ۲×۲

مثبت واقعی (TP): نتایج زمانی حاصل می‌شوند که مدل به درستی کلاس مثبت را پیش‌بینی می‌کند.

نمونه عضو دسته مثبت باشد و عضو همین کلاس تشخیص داده شود مثبت صحیح یا True Positive

منفی واقعی (TN): در جایی که مدل به درستی کلاس منفی را پیش‌بینی می‌کند، نتایج به دست می‌آیند.

نمونه عضو کلاس منفی باشد و عضو همین کلاس تشخیص داده شود منفی صحیح یا True Negative

مثبت کاذب (FP): که همچنین خطای نوع ۱ نیز نامیده می‌شود، نتیجه‌ای که در آن مدل به اشتباه کلاس مثبت را هنگامی که در واقع منفی است، پیش‌بینی می‌کند.

و در نهایت، نمونه عضو کلاس منفی باشد و عضو کلاس مثبت تشخیص داده شود مثبت کاذب یا False Positive

منفی کاذب (FN): که همچنین یک خطای نوع ۲ نیز نامیده می‌شود، نتیجه‌ای که در آن مدل به طور نادرست کلاس منفی را وقتی که واقعا مثبت است پیش‌بینی می‌کند.

نمونه عضو کلاس مثبت باشد و عضو کلاس منفی تشخیص داده شود منفی کاذب یا False Negative

(ب)

برای پروژه تشخیص مجرمان هک اسنپ فود، یک معیار ارزیابی مناسب می‌تواند معیار دقت (accuracy) باشد. معیار دقت نسبت تعداد پیش‌بینیهای صحیح به تعداد کل پیش‌بینیها را نشان میدهد. این معیار می‌تواند میزان کارایی مدل را در تشخیص مجرمان و بیگناهان اندازه‌گیری کند. اما این معیار تنها کافی نیست و باید با دیگر معیارهایی مانند صحت (precision)، بازخوانی (recall) و امتیاز F1 (F1-score) ترکیب شود. معیار صحت نسبت تعداد پیش‌بینیهای صحیح مثبت (مجرم) به تعداد کل پیش‌بینیهای مثبت را نشان میدهد. این معیار می‌تواند میزان اطمینان مدل را در تشخیص مجرمان نشان دهد. معیار بازخوانی نسبت تعداد پیش‌بینیهای صحیح مثبت به تعداد کل موارد مثبت واقعی (مجرم) را نشان میدهد. این معیار می‌تواند میزان توانایی مدل را در شناسایی مجرمان

نشان دهد. معیار امتیاز F1 میانگین هارمونیک معیارهای صحت و بازخوانی است. این معیار میتواند تعادل بین معیارهای صحت و بازخوانی را نشان دهد.

به طور خلاصه، برای ارزیابی عملکرد مدل در تشخیص مجرمان هک اسنپ فود، میتوان از معیار دقت به همراه معیارهای صحت، بازخوانی و امتیاز F1 استفاده کرد. این معیارها میتوانند نشان دهند که چقدر مدل موفق بوده است تا مجرمان و بیگناهان را از هم تمیز دهد و از اشتباه تشخیص ندادن افراد بیگناه به عنوان مجرم و تأمین امنیت مردم اطمینان حاصل کند.

این معیار میتواند میزان توانایی مدل را در شناسایی مجرمان نشان دهد. این معیار برای این پروژه مهم است چون ما میخواهیم از اشتباه تشخیص ندادن مجرمان و ایجاد خطر برای امنیت مردم جلوگیری کنیم. بنابراین، ما باید معیاری را انتخاب کنیم که حساسیت بالایی نسبت به موارد مثبت (مجرم) داشته باشد و تعداد موارد منفی اشتباه (false negative) را کمینه کند. معیار بازخوانی این ویژگی را دارد و میتواند مدل را به گونهای ارزیابی کند که احتمال دستگیری مجرمان را بالا ببرد و احتمال رهایی بیگناهان را کاهش ندهد.

پس بهترین انتخاب این است که FP را کمینه کنیم FP یعنی افراد بی گناه را اشتباه تشخیص ندهد که هکر هستند و FN را نیز که یعنی هکر ها هستند را اشتباه تشخیص ندهد که بیگناه هستند را هم باید کمینه کنیم.

سوال ۳

(الف)

تخمین چرخش وظیفه ای است که شامل پیش بینی زاویه چرخش یک تصویر است. این کار می تواند برای یک تکلیف یادگیری طبقه بندی و بازنمایی از چند جهت مفید باشد:

- تخمین چرخش می تواند به عنوان یک روش یادگیری خود نظارت برای یادگیری ویژگی های تصویر بدون استفاده از هیچ گونه حاشیه نویسی انسانی استفاده شود. با آموزش شبکه برای تشخیص چرخش یک تصویر، شبکه می تواند یاد بگیرد که ویژگی های معنایی و ثابت را از تصویر استخراج کند که می تواند برای کارهای پایین دستی مانند طبقه بندی مفید باشد. به عنوان مثال، Gidaris و همکاران. (۲۰۱۸) پیشنهاد استفاده از تخمین چرخش را به عنوان یک کار بهانه برای یادگیری بازنمایی تصویر ارائه کرد و در چندین معیار طبقه بندی تصویر به نتایج پیشرفته ای دست یافت.

- تخمین چرخش همچنین می تواند به عنوان یک تکنیک افزایش داده ها برای افزایش تنوع و استحکام داده های آموزشی برای یک کار طبقه بندی استفاده شود. با اعمال چرخش های تصادفی روی تصاویر و تنظیم برچسب ها بر اساس آن، شبکه می تواند دسته بندی تصاویر را در جهت های مختلف بیاموزد و خطر بیش از حد برازش را کاهش دهد. به عنوان مثال، وانگ و همکاران. (۲۰۱۹) پیشنهاد استفاده از تخمین چرخش به عنوان یک روش منظم سازی برای بهبود عملکرد شبکه های عصبی عمیق در وظایف طبقه بندی تصویر را ارائه کرد.

- تخمین چرخش همچنین می تواند به عنوان یک تکنیک استخراج ویژگی برای به دست آوردن نمایش های چرخشی ثابت از تصاویر برای یک کار طبقه بندی استفاده شود. با رمزگذاری اطلاعات چرخش تصاویر در یک بردار ویژگی، شبکه می تواند یاد بگیرد که تغییرات نامربوط تصاویر را نادیده بگیرد و بر ویژگی های اساسی کلاس ها تمرکز کند. به عنوان مثال، دنگ و همکاران. (۲۰۲۰)

پیشنهاد استفاده از تخمین چرخش را به عنوان روش یادگیری ویژگی برای به دست آوردن ویژگی‌های تغییرناپذیر چرخش برای طبقه‌بندی تصاویر ریزدانه پیشنهاد کرد.

به طور خلاصه، تخمین چرخش می‌تواند برای یک کار یادگیری طبقه‌بندی و نمایش به روش‌های مختلف مفید باشد، مانند یادگیری خود نظارت، افزایش داده‌ها، و استخراج ویژگی. با استفاده از تخمین چرخش، شبکه می‌تواند یاد بگیرد که ویژگی‌های معنایی و نامتغیر را از تصاویر استخراج کند و عملکرد خود را در وظایف طبقه‌بندی بهبود بخشد.

تخمین چرخش می‌تواند برای یادگیری بازنمایی در یک کار طبقه‌بندی از چند جهت مفید باشد:

یادگیری ویژگی‌های ثابت: با آموزش یک مدل بر روی تصاویر چرخانده، فرآیند یادگیری بازنمایی می‌تواند استخراج ویژگی‌هایی را که برای چرخش ثابت هستند تشویق کند. این می‌تواند منجر به نمایش ویژگی قوی‌تر شود که ویژگی‌های اساسی شی را علی‌رغم تغییر جهت‌گیری ثبت می‌کند.

تعمیم بهبود یافته: تخمین چرخش به عنوان بخشی از یادگیری بازنمایی می‌تواند منجر به فضای ویژگی شود که تعمیم بیشتری یافته و در جهت‌های مختلف اشیا قابل انتقال است. این می‌تواند توانایی مدل را برای طبقه‌بندی دقیق اشیاء حتی زمانی که با جهت‌گیری‌های نادیده در طول استنتاج ارائه می‌شود، افزایش دهد.

ویژگی‌های متنوع و آموزنده: آموزش تصاویر چرخانده، مدل را تشویق می‌کند تا ویژگی‌های متنوع و آموزنده‌ای را بیاموزد که ویژگی‌های شیء را از منظرهای متعدد ثبت می‌کند. در نتیجه، نمایش‌های آموخته شده می‌توانند درک غنی‌تر و جامع‌تری از اشیاء موجود در داده‌ها ارائه دهند.

افزایش داده‌ها: تخمین چرخش می‌تواند به عنوان شکلی از تقویت داده‌ها استفاده شود و به مدل اجازه می‌دهد تا از طیف وسیع تری از تغییرات ورودی یاد بگیرد. این داده‌های تقویت‌شده می‌تواند با قرار دادن مدل در معرض مجموعه‌ای جامع‌تر از تغییرات بصری مرتبط با شی گرای، به یادگیری بازنمایی بهتر منجر شود.

استحکام در دنیای واقعی: یادگیری بازنمایی با تخمین چرخش می‌تواند مدل را برای مدیریت بهتر سناریوهای دنیای واقعی که در آن اشیا در زوایای مختلف ظاهر می‌شوند آماده کند، بنابراین استحکام و سازگاری مدل را با محیط‌های متنوع و متغیر بهبود می‌بخشد.

در نتیجه، تخمین چرخش در یادگیری بازنمایی می‌تواند به ایجاد نمایش‌های ویژگی غیرمتغیر، تعمیم‌یافته و آموزنده کمک کند که توانایی مدل را برای انجام طبقه‌بندی دقیق در جهت‌های مختلف و زمینه‌های دنیای واقعی افزایش می‌دهد.

ب)

one-hot Vectors یا به انگلیسی **one-hot vectors**، نوعی از بازنمایی برداری از داده‌های گسسته هستند که در آن هر عنصر از داده‌ها توسط یک بردار دودویی نشان داده می‌شود که تنها یک درایه‌ی آن یک و بقیه صفر است. این روش برای جلوگیری از اینکه ماشین فرض کند که داده‌های گسسته دارای ترتیب یا ارزش عددی هستند، استفاده می‌شود. برای مثال، اگر ما بخواهیم چهار رنگ قرمز، آبی، سبز و زرد را بازنمایی کنیم، میتوانیم از این روش استفاده کنیم:

• قرمز = $[0, 0, 0, 1]$

• آبی = $[0, 1, 0, 0]$

• سبز = $[0, 0, 1, 0]$

• زرد = $[1, 0, 0, 0]$

این روش باعث می‌شود که ماشین نتواند فرض کند که رنگ قرمز بزرگتر یا کوچکتر از رنگ آبی است یا رنگ سبز با رنگ زرد قابل جمع یا تفریق است.

اما این روش هم دارای مشکلاتی است:

- این روش باعث می‌شود که فضای برداری بسیار بزرگ و پراکنده شود و نیاز به حافظه و پردازش زیادی داشته باشد.
 - این روش نمیتواند اطلاعات معنایی و تشابه بین داده‌ها را نشان دهد و تنها میتواند تفاوت بین داده‌ها را نشان دهد.
 - این روش ممکن است باعث کاهش عملکرد برخی از الگوریتم‌های یادگیری ماشین شود که به داده‌های پیوسته وجودی نیاز دارند.
- در پردازش زبان طبیعی، بردار تک داغ یک ماتریس $N \times 1$ (بردار) است که برای متمایز کردن هر کلمه در یک واژگان از هر کلمه دیگر در واژگان استفاده می‌شود. [۵] این بردار در تمام سلول‌ها به استثنای یک عدد ۱ در یک سلول که به طور منحصر به فرد برای شناسایی کلمه استفاده می‌شود، از ۰ تشکیل شده است. رمزگذاری یک‌طرفه تضمین می‌کند که یادگیری ماشینی فرض نمی‌کند که اعداد بالاتر مهم‌تر هستند. به عنوان مثال، مقدار "۸" بزرگتر از مقدار "۱" است، اما این باعث نمی‌شود که "۸" از "۱" مهمتر باشد. همین امر در مورد کلمات صادق است: ارزش "خنده دارتر" مهمتر از "خنده دار" نیست.
- رمزگذاری تک داغ در NLP برای رمزگذاری فاکتورهای دسته بندی به عنوان بردارهای باینری، مانند کلمات یا شناسه های بخشی از گفتار استفاده می شود.

این رویکرد مفید است زیرا الگوریتم‌های یادگیری ماشین معمولاً بر روی داده‌های عددی عمل می‌کنند، بنابراین نمایش داده‌های متنی به عنوان بردارهای عددی برای کارکرد این الگوریتم‌ها مورد نیاز است.

به عنوان مثال، در یک تکلیف تجزیه و تحلیل احساسات، ممکن است هر کلمه در یک جمله را به عنوان یک بردار رمزگذاری شده یک داغ توصیف کنیم و سپس از این بردارها به عنوان ورودی یک شبکه عصبی برای پیش بینی احساس جمله استفاده کنیم.

پیاده‌سازی آن نسبتاً ساده است و ما را قادر می‌سازد الگوریتم‌های یادگیری ماشینی را روی داده‌ها با ستون‌های طبقه‌بندی اعمال کنیم.

مشکل این است که ابعاد را افزایش می‌دهد بنابراین تمرین کندتر و پیچیده تر می‌شود. همچنین می‌تواند داده‌های پراکنده ایجاد کند زیرا بیشتر ورودی‌ها در ستون‌های جدید صفر خواهند بود. علاوه بر این، رمزگذاری یکباره فضای بیشتری را می‌گیرد اما هیچ اطلاعات جدیدی اضافه نمی‌کند زیرا فقط نمایش داده‌ها را تغییر می‌دهد.

حتی اگر دسته‌های نسبتاً کمی وجود داشته باشد، اگر داده‌ها دارای ردیف‌های زیادی باشند، رمزگذاری یکباره ممکن است این مشکلات را ایجاد کند.

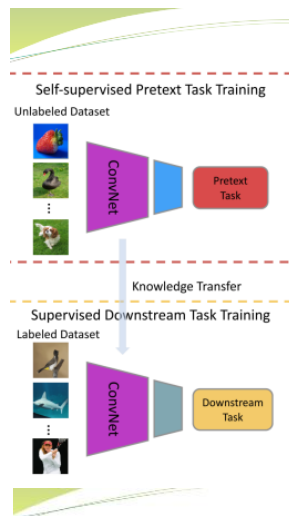
این می‌تواند منجر به افزایش ابعاد شود، زیرا یک ستون جداگانه برای هر دسته در متغیر ایجاد می‌شود. این می‌تواند مدل را پیچیده تر و کندتر کند.

این می‌تواند به داده‌های پراکنده منجر شود، زیرا اکثر مشاهدات در اکثر ستون‌های کدگذاری شده وان‌ها مقدار ۰ دارند. می‌تواند منجر به بیش از حد برازش شود، به خصوص اگر دسته بندی‌های زیادی در متغیر وجود داشته باشد و حجم نمونه نسبتاً کوچک باشد.

یک تکنیک قدرتمند برای پردازش داده‌های طبقه بندی شده است، اما می‌تواند منجر به افزایش ابعاد، پراکندگی و برازش بیش از حد شود. مهم است که با احتیاط از آن استفاده کنید و روش‌های دیگری مانند رمزگذاری ترتیبی یا رمزگذاری باینری را در نظر بگیرید.



طبق اسلاید های استاد :



یادگیری خودنظارتی (Self-Supervised)

- روش‌های یادگیری خود نظارتی برای یادگیری ویژگی‌های عمومی از داده‌های بدون برچسب در مقیاس بزرگ پیشنهاد شده‌اند
- مسئله‌های کمکی دو ویژگی مشترک دارند:
 - برای حل مسئله کمکی لازم است تا ویژگی‌های بصری توسط شبکه‌های کانولوشنی استخراج شوند
 - شبه‌برچسب‌های مورد نیاز در مسئله کمکی باید به صورت خودکار برای داده‌های مورد نظر قابل تولید باشند
- این رویکرد برای مسئله‌های غیر از بینایی کامپیوتر مانند پردازش صوت و پردازش متن نیز قابل استفاده است

Word2Vec

Context	Target
orange	

Skip-grams

- بجای اینکه همیشه از آخرین کلمات استفاده شود
 - به طور تصادفی یک کلمه را به عنوان Context انتخاب کنید
 - به طور تصادفی کلمه دیگری را در پنجره‌ای اطراف آن انتخاب کنید
- یک مسئله یادگیری با ناظر تنظیم کنید که در آن:
 - با توجه به کلمه Context
 - کلمه‌ای که به طور تصادفی انتخاب شده است را پیش‌بینی کنید

I want a glass of orange juice to go along with my cereal.

Word2Vec یک روش یادگیری بدون نظارت است که برای یافتن بازنمایی‌های برداری از کلمات استفاده می‌شود. این روش از دو مدل معماری مختلف برای یادگیری بازنماییها استفاده می‌کند: مدل کیسه کلمات پیوسته (Continuous Bag-of-Words) و مدل اسکپ گرام پیوسته (Continuous Skip-Gram). این دو مدل با استفاده از یک الگوریتم supervised-self مطابقت دارند.

الگوریتم supervised-self یک روش یادگیری نظارت شده است که از برچسب‌هایی که از داده‌های ورودی خودش تولید می‌کند، استفاده می‌کند. به عبارت دیگر، این روش از داده‌های ورودی به عنوان هم ورودی و هم خروجی استفاده می‌کند و سعی می‌کند یک تابع هدف را بیشینه کند که بر اساس شباهت بین ورودی و خروجی تعریف می‌شود.

مدل کیسه کلمات پیوسته یک مدل supervised-self است که از یک کلمه مرکزی به عنوان خروجی و چند کلمه اطراف آن به عنوان ورودی استفاده می‌کند. این مدل سعی می‌کند احتمال پیش‌بینی کلمه مرکزی را بر اساس کلمات اطراف آن بیشینه کند. مدل اسکپ گرام پیوسته یک مدل supervised-self است که از یک کلمه مرکزی به عنوان ورودی و چند کلمه اطراف آن به عنوان خروجی استفاده می‌کند. این مدل سعی می‌کند احتمال پیش‌بینی کلمات اطراف را بر اساس کلمه مرکزی بیشینه کند.

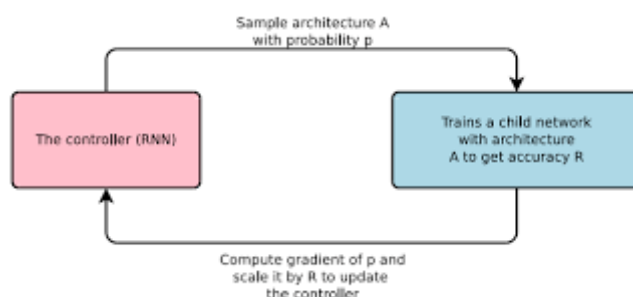
به طور خلاصه، Word2Vec با تعریف الگوریتمهای supervised-self مطابقت دارد چون از دادههای ورودی به عنوان هم ورودی و هم خروجی استفاده میکند و سعی میکند یک تابع هدف را بیشینه کند که بر اساس شباهت بین ورودی و خروجی تعریف میشود.

با این حال، قابلیت‌های یادگیری این مدل‌ها از زمان انتشار مقاله Word2Vec در سال ۲۰۱۳ که فضای NLP را متحول کرد، تکامل یافته است. ایده رویکردهای جاسازی کلمه ساده بود: به جای درخواست مدلی برای پیش‌بینی کلمه بعدی، می‌توانیم از آن بخواهیم کلمه بعدی را بر اساس زمینه قبلی پیش‌بینی کند.

Wav2Vec 2.0 از یک رویکرد آموزشی تحت نظارت خود برای تشخیص خودکار گفتار استفاده می‌کند که مبتنی بر ایده یادگیری متضاد است. یادگیری نمایش گفتار در یک مجموعه داده عظیم خام (بدون برچسب) میزان داده‌های برچسب گذاری شده مورد نیاز برای دریافت نتایج رضایت بخش را کاهش می‌دهد.

سوال ۴

الف



طراحی اولیه (Zoph & Le 2017) NAS شامل یک کنترل کننده مبتنی بر RL برای پیشنهاد معماری مدل کودک برای ارزیابی است. کنترل کننده به عنوان یک RNN پیاده‌سازی می‌شود و یک دنباله با طول متغیر از توکن‌های مورد استفاده برای پیکربندی یک معماری شبکه را خروجی می‌دهد.

فضای عمل: فضای عمل فهرستی از نشانه‌ها برای تعریف شبکه فرزند پیش‌بینی شده توسط کنترل کننده است (بیشتر را در بخش بالا ببینید). کنترلر عمل "a1:T" را خروجی می‌دهد، که در آن T تعداد کل توکن‌ها است.

پاداش: دقت شبکه کودک که می‌تواند در همگرایی به دست آید، پاداش آموزش کنترل کننده، R.

Loss: NAS پارامترهای کنترلر تتا را با از دست دادن REINFORCE بهینه می‌کند. ما می‌خواهیم پاداش مورد انتظار (دقت بالا) را با گرادینان به صورت زیر به حداکثر برسانیم. نکته خوب در اینجا با گرادینان سیاست این است که حتی زمانی که پاداش غیرقابل تمایز باشد، کار می‌کند.

$$\nabla_{\theta} J(\theta) = \sum_{t=1}^T \mathbb{E}[\nabla_{\theta} \log P(a_t | a_{1:(t-1)}; \theta) R]$$

MetaQNN یک عامل را آموزش می دهد تا به طور متوالی لایه های CNN را با استفاده از یادگیری Q با استراتژی اکتشاف حریصانه اپسیلون و بازپخش تجربه انتخاب کند. پاداش نیز دقت اعتبارسنجی است.

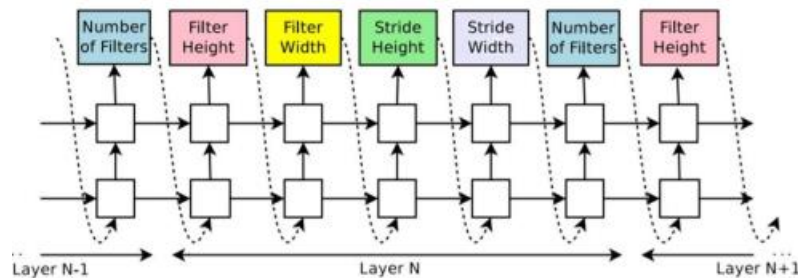
$$Q^{(t+1)}(s_t, a_t) = (1 - \alpha)Q^{(t)}(s_t, a_t) + \alpha(R_t + \gamma \max_{a \in \mathcal{A}} Q^{(t)}(s_{t+1}, a'))$$

که در آن حالت s_t یک عملیات لایه و پارامترهای مرتبط با آن است. یک عمل a ارتباط بین عملیات را تعیین می کند. Q-value متناسب با میزان اطمینان ما در دو عملیات متصل است که منجر به دقت بالا می شود.

برای به روزرسانی پارامترهای کنترلر از یادگیری تقویتی استفاده می شود تا دقت اعتبارسنجی مورد انتظار را بیشینه کند.

یادگیری تقویتی یک رویکرد یادگیری ماشینی است که شامل یادگیری از آزمون و خطا از طریق تعامل با یک محیط است. یک عامل یادگیری تقویتی یاد می گیرد که اقداماتی را انجام دهد که سیگنال پاداش را به حداکثر می رساند، که نشان دهنده مطلوبیت وضعیت agent است. یادگیری تقویتی را می توان برای جستجوی خودکار ساختار شبکه یا ابرهای شبکه با تعریف یک تابع پاداش مناسب و فضای جستجو برای معماری های شبکه استفاده کرد.

با استفاده از رویکرد RL، یک شبکه عصبی بازگشتی را برای تولید معماری های کانولوشن آموزش دهید. شکل زیر یک کنترل کننده RNN را نشان می دهد که به طور متوالی پارامترهای CNN مرتبط با لایه های کانولوشن را تولید می کند.



هر خروجی دنباله ای توسط یک کلاسیر softmax پیش بینی می شود و سپس به عنوان ورودی دنباله بعدی استفاده می شود. مجموعه پارامترها شامل ارتفاع و عرض فیلتر، ارتفاع و عرض گام و تعداد فیلتر در هر لایه است. طراحی یک معماری زمانی پایان می یابد که تعداد لایه ها به مقدار از پیش تعریف شده ای برسد که در طول آموزش افزایش می یابد. دقت معماری طراحی شده به عنوان پاداشی برای آموزش کنترل کننده RNN با RL به منظور به حداکثر رساندن دقت اعتبارسنجی مورد انتظار معماری های بعدی تغذیه می شود.

تعامل عامل-محیط: در NAS، عامل با محیطی تعامل دارد که در آن حالت یک معماری شبکه عصبی خاص را نشان می دهد و اقدامات مربوط به تغییرات معماری (به عنوان مثال، افزودن یا حذف لایه ها، تنظیم اتصالات) است.

۱. سیگنال پاداش: عامل RL یک سیگنال پاداش را بر اساس عملکرد معماری شبکه عصبی در یک کار خاص دریافت می کند (به عنوان مثال، دقت در مجموعه اعتبار سنجی). پاداش به عنوان معیاری برای کیفیت معماری عمل می کند و روند جستجو را هدایت می کند.
۲. یادگیری خط مشی: عامل RL خط مشی را می آموزد که وضعیت های معماری شبکه را به اقداماتی ترسیم می کند، به گونه ای که پاداش تجمعی مورد انتظار را در طول زمان به حداکثر می رساند. این عامل را قادر می سازد تا به طور مکرر پیکربندی های معماری را کاوش و بهره برداری کند.
۳. کاوش و بهره برداری: RL اجازه می دهد تا تعادلی بین کاوش طرح های جدید معماری و بهره برداری از پیکربندی های نویدبخش شناخته شده برقرار شود، در نتیجه جستجوی معماری های شبکه با کارایی بالا را تسهیل می کند.
۴. آموزش و ارزیابی مدل: در طول فرآیند NAS، RL می تواند برای آموزش و ارزیابی معماری های نمونه مورد استفاده قرار گیرد و به تدریج کیفیت معماری های کشف شده را از طریق یادگیری تقویتی بهبود بخشد.
۵. کارایی نمونه: الگوریتم های RL را می توان برای بهینه سازی کارایی نمونه طراحی کرد، به این معنی که می توانند به طور موثر فضای معماری را با منابع محاسباتی محدود کاوش کنند.

یکی از راه های استفاده از یادگیری تقویتی برای جستجوی ساختار شبکه، استفاده از شبکه عصبی بازگشتی (RNN) به عنوان کنترل کننده ای است که معماری های شبکه را به طور متوالی تولید می کند. کنترل کننده با روش های گرایان سیاست برای بهینه سازی پاداش مورد انتظار، که معمولاً دقت اعتبار سنجی شبکه های تولید شده است، آموزش دیده است. کنترل کننده می تواند با نمونه برداری از یک فضای جستجوی از پیش تعریف شده، معماری های مختلف شبکه را کشف کند، مانند تعداد و نوع لایه ها، توابع فعال سازی، فرآپارامترها و غیره. همچنین کنترل کننده می تواند با انتخاب معماری های شبکه که احتمال بالایی دارند، از دانش آموخته شده بهره برداری کند. از دستیابی به پاداش های بالا به این ترتیب، عامل یادگیری تقویتی می تواند یاد بگیرد که معماری های شبکه ای را که برای یک کار آموزشی مناسب هستند، ایجاد کند.

راه دیگر برای استفاده از یادگیری تقویتی برای جستجوی ساختار شبکه، استفاده از یک شبکه عصبی کانولوشن (CNN) به عنوان یک تابع مقدار است که امتیازات معماری شبکه را تقریب می کند. تابع ارزش با روش های یادگیری Q آموزش داده می شود تا پاداش مورد انتظار هر معماری شبکه را تخمین بزند، که همچنین دقت اعتبار سنجی شبکه ها است. تابع مقدار می تواند معماری های مختلف شبکه را با اعمال اغتشاشات تصادفی در معماری شبکه فعلی، مانند افزودن، حذف یا تغییر لایه ها، بررسی کند. تابع ارزش همچنین می تواند از دانش آموخته شده با انتخاب معماری شبکه ای که بالاترین پاداش تخمینی را دارد، بهره برداری کند. به این ترتیب، عامل یادگیری تقویتی می تواند ارزیابی و بهبود معماری شبکه را به طور مکرر یاد بگیرد.

ب

بیکر و همکاران MetaQNN، یک الگوریتم فرامدل سازی را معرفی کرد که از یادگیری تقویتی برای تولید خودکار معماری های CNN با کارایی بالا در حوزه طبقه بندی تصویر [BGNNR17] استفاده می کند. این الگوریتم وظیفه استفاده از یادگیری Q در آموزش یک عامل را در انتخاب متوالی لایه های شبکه عصبی و فرآپارامترهای آنها در نظر می گیرد. الگوریتم MetaQNN هنگام

ر سیدن به یک حالت خاتمه (اعم از لایه Softmax یا میانگین جهانی)، عملکرد شبکه عصبی تولید شده را ارزیابی می کند و با استفاده از دقت به عنوان پاداش، از Q-Learning برای تنظیم پاداش مورد انتظار انتخاب های انجام شده در طول عصبی استفاده می کند.

MetaQNN یک عامل را آموزش می دهد تا به طور متوالی لایه های CNN را با استفاده از یادگیری Q با استراتژی اکتشاف حریصانه اپسیلون و بازپخش تجربه انتخاب کند. پاداش نیز دقت اعتبارسنجی است.

$$Q^{(t+1)}(s_t, a_t) = (1 - \alpha)Q^{(t)}(s_t, a_t) + \alpha(R_t + \gamma \max_{a \in A} Q^{(t)}(s_{t+1}, a'))$$

که در آن حالت s_t یک عملیات لایه و پارامترهای مرتبط با آن است. یک عمل a ارتباط بین عملیات را تعیین می کند. Q-value متناسب با میزان اطمینان ما در دو عملیات متصل است که منجر به دقت بالا می شود.

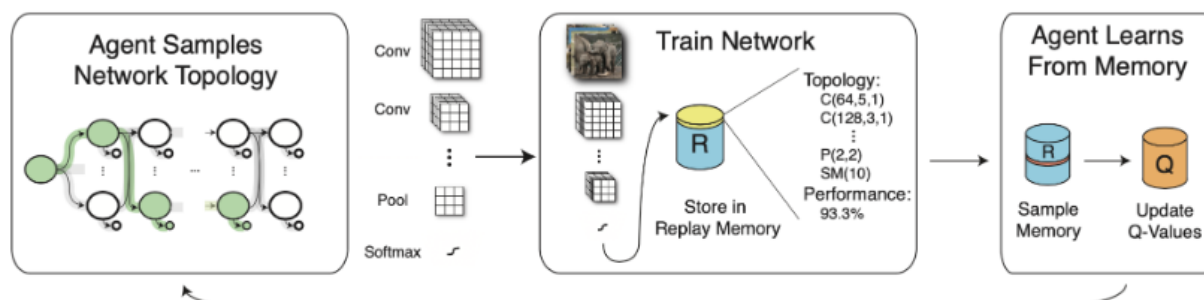


Fig. 9. Overview of MetaQNN - designing CNN models with Q-Learning.
(Image source: Baker et al. 2017)

روش های مبتنی بر یادگیری تقویتی (RL) در واقع می توانند برای جستجوی فراپارامترهای بهینه، از جمله اندازه تصویر ورودی و تعداد لایه ها در معماری شبکه های عصبی مورد استفاده قرار گیرند. در اینجا نحوه استفاده از RL برای جستجوی این هایپرپارامترها آمده است:

۱. اندازه تصویر ورودی

RL: را می توان برای تعیین اندازه تصویر ورودی بهینه برای یک کار یا معماری مدل مورد استفاده قرار داد.

عامل می تواند اندازه های مختلف تصویر را به عنوان اقدامات انتخاب کند و بر اساس عملکرد مدل در هنگام آموزش و ارزیابی با استفاده از تصاویر آن اندازه ها، جایزه دریافت کند.

عامل RL می تواند طیف وسیعی از اندازه های تصویر را کاوش کند و یاد بگیرد که اندازه ای را انتخاب کند که معیار عملکرد را به حداکثر می رساند، با در نظر گرفتن مبادلات بین هزینه محاسباتی، نمایش ویژگی و دقت مدل.

اندازه تصویر ورودی می‌تواند تأثیر قابل توجهی بر عملکرد مدل داشته باشد، به‌ویژه در کارهایی مانند تشخیص و تقسیم‌بندی اشیاء، جایی که بر میدان دریافتی و سطح جزئیات گرفته‌شده توسط مدل تأثیر می‌گذارد.

۲. تعداد لایه‌ها:

RL می‌تواند تعداد بهینه لایه‌ها را در یک معماری عصبی با در نظر گرفتن تعداد لایه‌ها به عنوان اقدامات و یادگیری بهترین پیکربندی معماری از طریق آزمون و خطا جستجو کند.

عامل RL می‌تواند یک فضای گسسته یا پیوسته از تعداد لایه‌ها را هدایت کند و عملکرد معماری‌ها را با پیچیدگی‌های لایه‌های مختلف ارزیابی کند.

انتخاب تعداد لایه‌ها می‌تواند بر ظرفیت مدل، قدرت بازنمایی و توانایی در گرفتن ویژگی‌های سلسله‌مراتبی تأثیر بگذارد. RL می‌تواند جستجو برای تعادل بهینه بین پیچیدگی مدل و تعمیم را هدایت کند.

۳. دلایل جستجوی مبتنی بر RL:

فضای جستجوی پیچیده: هم اندازه تصویر ورودی و هم تعداد لایه‌ها نشان‌دهنده هاپارامترهای پیچیده با طیف وسیعی از مقادیر ممکن است. RL می‌تواند به طور موثر این فضای جستجوی با ابعاد بالا و غیرخطی را کاوش کند.

روابط غیر پیش پا افتاده: تأثیر این فراپارامترها بر عملکرد مدل بی‌اهمیت است و مقادیر بهینه آنها ممکن است به مجموعه داده و وظیفه خاص بستگی داشته باشد. RL می‌تواند یاد بگیرد که این روابط را از طریق اکتشاف و بهره‌برداری تکراری هدایت کند.

کارایی محاسباتی: رویکردهای RL را می‌توان برای بهینه‌سازی فرآیند جستجو برای این ابرپارامترها، با در نظر گرفتن کارایی محاسباتی و محدودیت‌های منابع، طراحی کرد.

سازگاری: جستجوی مبتنی بر RL سازگاری را با تغییر توزیع داده‌ها، نیازمندی‌های کار، و محدودیت‌های سخت‌افزاری امکان‌پذیر می‌کند و امکان کشف پیکربندی‌های فراپارامتر را فراهم می‌کند که در زمینه‌های مختلف مؤثر هستند.

امکان استفاده از رویکرد جستجوی یادگیری تقویتی برای اندازه تصویر ورودی و تعداد لایه‌ها وجود دارد. این دو ابرپارامتر می‌توانند تأثیر زیادی بر کارایی مدل تشخیص اشیاء داشته باشند. اندازه تصویر ورودی می‌تواند بر روی کیفیت و سرعت تشخیص اشیاء تأثیر گذار باشد. اگر اندازه تصویر ورودی بزرگ باشد، ممکن است دقت تشخیص اشیاء بالاتر باشد، اما سرعت آن کمتر باشد. اگر اندازه تصویر ورودی کوچک باشد، ممکن است سرعت تشخیص اشیاء بالاتر باشد، اما دقت آن کمتر باشد. بنابراین، باید یک تعادل بین اندازه تصویر ورودی و کارایی مدل یافت. تعداد لایه‌ها می‌تواند بر روی توانایی مدل در یادگیری ویژگی‌های مرتبط با اشیاء تأثیر گذار باشد. اگر تعداد لایه‌ها زیاد باشد، ممکن است مدل بتواند ویژگی‌های پیچیده‌تری را یاد بگیرد، اما خطر بیش‌برازش و افزایش پیچیدگی محاسباتی وجود داشته باشد. اگر تعداد لایه‌ها کم باشد، ممکن است مدل ساده‌تر و سریع‌تر باشد، اما خطر کم‌برازش و کاهش قدرت تمایز وجود داشته باشد. بنابراین، باید یک تعادل بین تعداد لایه‌ها و کارایی مدل یافت.

به طور خلاصه، رویکرد جستجوی یادگیری تقویتی میتواند برای بهینه‌سازی ابر پارامترها و پارامترهای معماری شبکه‌های عصبی برای مسئله تشخیص اشیا استفاده شود. این رویکرد میتواند برای اندازه تصویر ورودی و تعداد لایه‌ها نیز کاربرد داشته باشد. این دو ابر پارامتر میتوانند تاثیر زیادی بر کارایی مدل تشخیص اشیا داشته باشند و باید یک تعادل بین آنها و کارایی مدل یافت شود.

با استفاده از یادگیری تقویتی برای جستجوی فراپارامترهایی مانند اندازه تصویر ورودی و تعداد لایه‌ها، محققان می‌توانند فرآیند شناسایی پیکربندی‌هایی را که منجر به عملکرد بهینه مدل می‌شوند، به طور خودکار انجام دهند، با نیازهای مختلف سازگار شوند و فضای ابرپارامتر را به طور کارآمد بررسی کنند.

سوال ۵

شبکه های مولد تخصصی (GAN) از دو شبکه عصبی تشکیل شده‌اند: یک مولد (generator) و یک متمایزکننده (discriminator). مولد سعی دارد تصاویر جدیدی را تولید کند که شبیه به داده‌های واقعی باشند و متمایزکننده سعی دارد تصاویر واقعی و تولید شده را از هم تفکیک کند. این دو شبکه در یک بازی تخصصی با هم رقابت میکنند و هر کدام سعی میکنند عملکرد خود را بهبود بخشند.

تابع ضرر مولد و متمایزکننده معیاری برای اندازه‌گیری عملکرد این دو شبکه است. تابع ضرر مولد میزان توانایی مولد را در فریب دادن متمایزکننده نشان میدهد و تابع ضرر متمایزکننده میزان توانایی متمایزکننده را در تشخیص صحیح تصاویر نشان میدهد. این دو تابع ضرر به صورت متناقض با هم هستند، یعنی کاهش یکی منجر به افزایش دیگری میشود.

اما اینکه مقدار تابع ضرر مولد و متمایزکننده در پایان epoch اول و ۱۰۰ تقریباً یکسان باشد، به این معنی نیست که کیفیت تصاویر تولید شده در این دو epoch هم یکسان باشد. چرا که مقدار تابع ضرر تنها نشان دهندهی نسبت قدرت مولد و متمایزکننده است، نه کیفیت تصاویر خود. ممکن است در epoch اول، هر دو شبکه ضعیف باشند و تصاویر تولید شده کیفیت پایینی داشته باشند، اما در epoch ۱۰۰، هر دو شبکه قوی شده باشند و تصاویر تولید شده کیفیت بالایی داشته باشند. در این صورت، مقدار تابع ضرر ممکن است تغییر چندانی نکند، اما کیفیت تصاویر به طور قابل توجهی افزایش یابد.

به طور خلاصه، مقدار تابع ضرر مولد و متمایزکننده در شبکه GAN تنها نشان دهندهی نسبت قدرت این دو شبکه است و نمیتوان از آن به عنوان معیاری برای ارزیابی کیفیت تصاویر تولید شده استفاده کرد. برای ارزیابی کیفیت تصاویر تولید شده، میتوان از معیارهای دیگری مانند امتیاز Inception (Inception Score) یا امتیاز FID (Fréchet Inception Distance) استفاده کرد.

توضیحات بیش تر : GPT

اگر توابع زیان مولد و بحرانی در پایان دوره اول و در دوره ۱۰۰ تقریباً یکسان باشد، نشان می دهد که GAN به نقطه ای رسیده است که مولد در حال ایجاد تصاویری است که تشخیص آنها از واقعی برای متمایز کننده بسیار دشوار است. تصاویر. با این حال، کیفیت تصاویر تولید شده در دوره اول و دوره ۱۰۰ ممکن است به دلایل مختلفی از جمله:

فروپاشی حالت: GAN ممکن است از فروپاشی حالت رنج ببرد، که در آن ژنراتور انواع محدودی از تصاویر را تولید می کند که منجر به عدم تنوع در تصاویر تولید شده می شود.

بی ثباتی تمرین: GANها به دلیل دشواری در آموزش بدنام هستند و می توانند از بی ثباتی تمرینی رنج ببرند. علیرغم داشتن مقادیر تلفات مشابه، فرآیند آموزش در مراحل اولیه و مراحل بعدی ممکن است دینامیک متفاوتی را نشان دهد که به طور بالقوه بر کیفیت تصاویر تولید شده تأثیر می گذارد.

Overfitting یا Underfitting: GAN ممکن است داده های آموزشی را بیش از حد برازش کند یا کمتر از آن استفاده کند و در نتیجه کیفیت تصویر بین دوره ها تفاوت داشته باشد.

نویز و تصادفی بودن: فرآیند آموزش ممکن است شامل تصادفی و نویز باشد که منجر به تغییرات در تصاویر تولید شده حتی زمانی که عملکردهای از دست دادن مشابه هستند.

دینامیک یادگیری: پویایی یادگیری GAN ممکن است در طول زمان تغییر کند. حتی اگر توابع از دست دادن مشابه باشند، شبکه ممکن است الگوها و ویژگی های جدیدی را در دوره صدم یاد گرفته باشد که منجر به بهبود کیفیت تصویر می شود.

فراپارامترها و تغییرات معماری: ابرپارامترها یا معماری مدل ممکن است در طول آموزش تغییر کرده باشند، یا ممکن است استراتژی های آموزشی متفاوتی مانند زمان بندی نرخ یادگیری یا تکنیک های منظم سازی اعمال شده باشند.

تغییرات مجموعه داده: GAN ممکن است در طول دوره آموزشی با توزیع داده ها یا تغییرات متفاوتی در داده های آموزشی مواجه شده باشد که بر کیفیت تصاویر تولید شده تأثیر می گذارد.

پیچیدگی ذاتی: مدل های یادگیری عمیق، از جمله GANها، پیچیدگی بالایی از خود نشان می دهند و رابطه بین توابع از دست دادن و کیفیت تصویر همیشه ساده نیست، به ویژه در حضور بهینه سازی غیرمحدب، گرادیان های ناپدید و سایر چالش ها.

به طور خلاصه، در حالی که توابع از دست دادن بینش های ارزشمندی را در مورد آموزش GAN ها ارائه می دهند، ممکن است ظرافت های فرآیند تولید تصویر را به طور کامل نشان ندهند، و عوامل دیگر در کیفیت تصاویر تولید شده نقش دارند.