DOI:https://doi.org/10.22034/sdsb.1.4.44

توسعه پایدار و ساختمان هوشمند

https://sdsb.buqaen.ac.ir

جلد ۱، شماره ۴، زمستان ۱۴۰۴

بررسی انواع لایه‌های ادغام برای طبقه‌بندی تصاویر مبتنی بر یادگیری عمیق

اعظم نوروزی[1*]

۱- دانشکده مهندسی- گروه عمران و معماری- دانشگاه تربت حیدریه- تربت حیدریه- ایران.

چکیده

لایه ادغام به طور گسترده‌ای برای طبقه‌بندی تصاویر با بیش از یک دهه یادگیری عمیق استفاده شده است. بسیاری از مدل‌ها با استفاده از لایه ادغام، سعی کردند بر معایب الگوریتم سنتی غلبه کنند که این امر به ویژگی‌های طراحی دستی بستگی دارد. به طور کلّی، مدل‌های یادگیری عمیق، اغلب حاوی پارامترهای قابل آموزش هستند و برای دست‌یابی به عملکرد بهینه به نمونه‌های برچسب‌گذاری‌شده زیادی نیاز دارند. این مقاله بر روی پارامترهای بهینه و کارهای قبلی مربوطه تمرکز کرده است. رویکردهای پیشرفته مختلف از طریق چندین آزمایش مورد بررسی قرار گرفته‌اند. نتایج ارزیابی بر روی مجموعه داده ۱۰-□□□□□ نشان داد که سریع‌ترین و بهترین روش، ساده‌ترین روش از نظر محاسباتی می‌باشد که روش ادغام ماکسیمم است. پس از روش ماکسیمم، روش ادغام □□□ بهترین نتایج را دارد، اما زمان تشخیص را تقریباً پنجاه درصد افزایش داده است. این افزایش زمان به دلیل پیچیدگی‌های محاسباتی لگاریتمی و نمایی آن است. روش ترکیبی در آزمایش‌ها نمی‌توانست بهتر از ادغام ماکسیمم، در مورد معیارهای طبقه‌بندی رفتار کند. اما این روش، کمترین مقدار خطا را در مرحله آموزش دارد.

اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۴/۱۰/۲۱

تاریخ پذیرش: ۱۴۰۴/۱۲/۰۶

واژه‌های کلیدی:

یادگیری عمیق، لایه ادغام، طبقه بندی تصویر،شبکه عصبی کانولوشن، یادگیری ماشین.

* نویسنده مسئول، پست الکترونیکی: a.noroozi@torbath.ac.ir

Journal of Sustainable Development and Smart Building, vol. 1, no. 4

Digital Object Identifier https://doi.org/10.22034/sdsb.1.4.44

۱. مقدّمه

۱-۱. پیشینه

طبقه‌بندی تصویر، یک فرایند پیچیده است که عوامل زیادی را در نظر می گیرد [۱، ۲]. این روش‌ها، ویژگی‌هایی را با سطح بازنمایی بالاتر توسعه می‌دهند و موارد نامربوط را حذف می‌کنند. علاوه بر این، یک تصویر در طول فرایند طبقه‌بندی تصویر از لایه‌های انتقال مختلف عبور می‌کند تا به خروجی برسد [۳، ۴]. اخیراً تکنیک‌های جدیدی در حوزه طبقه‌بندی تصاویر ارائه شده است تا محققان بتوانند طبقه‌بندی آن‌ها را بر اساس معیارهایی مانند پیچیدگی زمانی و دقت ارزیابی کنند. انگیزه اولیه برای ارزیابی روش‌های طبقه‌بندی تصاویر، انتخاب یک رویکرد مناسب با یک فرایند کامل است [۵]. تکنیک‌های مختلف ممکن است برای انجام وظایف مختلف، به طور متفاوت استفاده شود. تعیین معیارهای ارزیابی، مانند اندازه و نوع داده، برای اعتبارسنجی، بهینه‌ترین و ضروری‌ترین راه برای انجام یک کار خاص و نتیجه مورد انتظار است [۶، ۷].

۱-۲. خلأهای تحقیقاتی

در طول دهه گذشته، یادگیری ماشین در زمینه‌های مختلف، محبوبیت پیدا کرده است و مفهوم محاسبات عصبی را تغییر داده است [۸، ۹]. بر خلاف رویکردهای غیر یادگیری که بر قوانین خاص دامنه تکیه نمی‌کنند، روش‌های یادگیری ماشینی، دانش را از طریق تحلیل داده‌ها به دست می‌آورند. این بدان معناست که روش‌ها به شدت به نحوه ساختار و ارائه داده‌ها بستگی دارد که به تخصص جامع، طراحی ظریف و داده‌های خام نیاز دارد [۵].

به طور کلّی، تکنیک‌های یادگیری ماشین به صورت زیر طبقه‌بندی[۱] می‌شوند [۱۰، ۱۱]: روش‌های یادگیری تحت نظارت[۲][۱۲، ۱۳] که مجموعه داده‌ها را بر اساس مدل‌ها آموزش می‌دهد. انواع اصلی این روش‌ها طبقه‌بندی و رگرسیون[۳]است. الگوریتم‌های یادگیری بدون نظارت[۴][۱۴]، سعی می‌کنند شباهت بین داده‌های ورودی را بیابند. خوشه‌بندی،[۵]روش مناسب برای این دسته است. یادگیری نیمه نظارت شده[۶][۱۵، ۱۶] از داده‌های برچسب‌دار و بدون برچسب استفاده می‌کند. این تکنیک، یک نوع از مدل یادگیری نظارت شده است. یادگیری تقویتی تحت نظارت[۷][۱۷، ۱۸]، یک مدل مبتنی بر روانشناسی است. این روش، زمانی استفاده می‌شود که اطلاعاتی در مورد مشکل وجود نداشته باشد. یادگیری تکاملی [۱۹]، شامل توسعه بیولوژیکی است که در آن موجودات بیولوژیکی بررسی و اصلاح می‌شود تا تداوم خود را توسعه دهد. رویکردهای یادگیری عمیق از طریق لایه‌های پردازشی مختلف انجام می‌شود. امروزه، یادگیری عمیق، یک موضوع تحقیقاتی مرسوم است [۲]، و شبکه عصبی کانولوشن[۸]به دلیل دقّت بالا، یک مدل پیشرو برای خوشه-بندی، تشخیص و طبقه‌بندی تصاویر است [۶، ۲۰]. روش‌های یادگیری عمیق با ترکیب بخش‌های[۹]ساده و غیر خطی، نمایش یک سطحی را به سطح بالاتر تبدیل می‌کند. در نتیجه، این تبدیل-های ترکیبی، برخی از توابع پیچیده را آموزش می‌دهد. چالش‌هایی در طبقه‌بندی تصاویر مبتنی بر یادگیری عمیق وجود دارد که عبارت است از [۲۱]:

- در نظر گرفتن تعداد ورودی‌ها و تعیین ستون‌های غیر کمک-کننده،
- تعیین تعداد نورون‌ها در تمام لایه‌های پنهان،
- تعیین تعداد لایه‌های پنهان،[۱۰]
- توابع فعال‌سازی،[۱۱]
- روش‌های بهینه‌سازی،
- تابع محوسازی،[۱۲]
- تعیین تعداد دوره ها[۱۳]و اندازه دسته[۱۴]در مرحله پردازش.

حوزه یادگیری عمیق، نیازمند پیشرفت مستمر برای غلبه بر مشکلات است. اگر آن‌ها هر روش یادگیری عمیق را انجام دهند، ناگزیر به چالش‌های نظری و تجربی منجر می‌شود [۲۲]. شبکه‌های عصبی عمیق، توجّه تحقیقاتی گسترده‌ای را به خود جلب کرده است، بنابراین، مدل‌های مختلفی را برای کاربردهای مختلف پیشنهاد کرده‌اند [۲۳]. مدل شبکه عصبی کانولوشن، ابعاد نقشه[۱۵]ویژگی را کاهش می‌دهد و در نتیجه، پیچیدگی زمانی کمتری دارد [۲۴، ۲۵].

۱-۳. اهداف

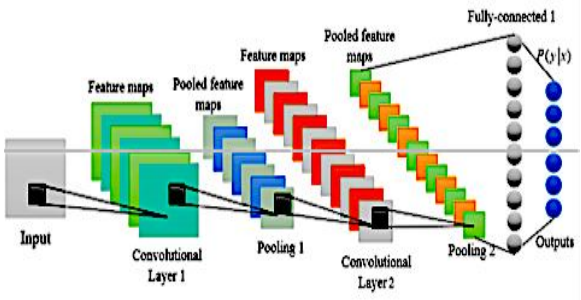
1. Classification.
2Supervised Learning.
3Regression.
4Unsupervised Learning.
5Clustering.
6Semi-supervised Learning.
7Reinforcement Learning.
8Convolution Neural Network.
9Module.
10Hidden Layer.
11Activation Function.
12Decay Function.
13Epoch.
14Batch size.
15Feature map.

این مقاله، یک بررسی کامل و جامع از انواع لایه‌های پرکابرد ادغام[۱] پیشنهاد شده در روش‌های یادگیری عمیق در طبقه‌بندی تصاویر را انجام می‌دهد. ابتدا محبوب‌ترین لایه‌های ادغام شامل میانگین،[۲] ماکسیمم،[۳] ترکیبی،[۴] Lp، ادغام مبتنی بر اهمیت محلّی،[۵] تصادفی،[۶] تابع نرم[۷] و DWT[۸] در نظر گرفته شد. سپس، آزمایش‌های زیادی با رویکردهای پیشرفته مختلف انجام شده است و نتایج برای نشان‌دادن جهت بالقوه تحقیق، مورد بررسی و ارزیابی قرار گرفته شده است.

ادامه مقاله، شامل بخش‌های زیر است: بخش ۲، مدل شبکه عصبی کانولوشن را توضیح می‌دهد، بخش ۳، بر معرفی لایه ادغام و مدل‌های محبوب آن تمرکز دارد، بخش ۴، نتایج ارزیابی را ارائه می‌دهد و بخش ۵، نتیجه بررسی را بیان می‌کند.

۲. شبکه عصبی کانولوشن

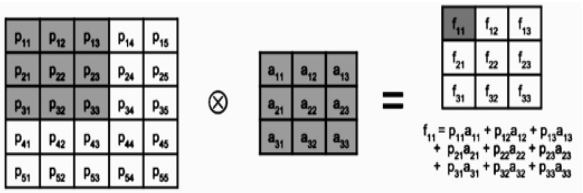
این شبکه نوع [23] شبکه عصبی عمیق است که معمولاً برای طبقه‌بندی و بخش‌بندی اشیاء[۹] استفاده می‌شود [۲۶]. قدرت لایه‌های متعدد آموزش، شبکه‌های عصبی کانولوشن را به بخش مهمی از روش‌های یادگیری عمیق تبدیل کرده است. شبکه‌های عصبی کانولوشن از لایه‌های مختلفی تشکیل شده‌اند که هر یک وظایف خاصی را انجام می‌دهند. گام اول آموزش[۱۰] در این شبکه‌ها، مرحله پیش انتشار[۱۱] شناخته می‌شود که در آن داده‌های ورودی به شبکه ارسال می‌شود. این عملیات، نقاط بین نورون ورودی و پارامترها را ضرب می‌کند و عملیات کانولوشن را در لایه ها انجام می‌دهد. سپس خروجی شبکه، محاسبه می‌شود. پارامترهای آموزش شبکه با استفاده از خطای شبکه تنظیم می‌شود، سپس با استفاده از تعداد خطاهای محاسبه شده، مرحله پس انتشار[۱۲] آغاز می‌شود. مرحله آموزش تا زمانی که میزان خطا[۱۳] به یک آستانه[۱۴] برسد، تکرار می‌شود. شکل ۱ یک معماری مدل عصبی کانولوشن را نشان می‌دهد. این مدل، دارای سه لایه اصلی به شرح زیر است [۲۷]:



شکل ۱: معماری شبکه عصبی.

۲-۱. لایه کانولوشن

این لایه از چندین فیلتر تشکیل شده است که با ورودی شبکه، تعامل دارد و نقشه‌های ویژگی را ایجاد می‌کند. فرایند یادگیری شبکه عصبی عمیق، به‌روزرسانی مداوم عناصر فیلتر است. به عبارت دیگر، وزن شبکه‌های عصبی، ضرایب این فیلترها هستند. بنابراین، لایه کانولوشن، چندین نقشه ویژگی ایجاد می‌کند که هر کدام تصویر حاصل از ضرب تصویر اصلی در یکی از فیلترها است. ابعاد خروجی، کمتر از ابعاد ورودی در اجرای فرایند کانولوشن است. با توجّه به اینکه این لایه با استفاده از روش‌هایی مانند گام حرکت[۱۵] یا لبه‌گذاری[۱۶] ویژگی را استخراج می‌کند، ابعاد خروجی و ورودی، برابر می‌شود و لایه ادغام ابعاد را کاهش می‌دهد. شکل ۲ یک مثال فرایند کانولوشن را نشان می‌دهد.



شکل ۲: فرایند کانولوشن.

۲-۲. لایه ادغام

این لایه، بخشی ضروری از سیستم‌های مبتنی بر کانولوشن است [۲۹] که ابعاد ویژگی‌ها را کاهش می‌دهد، در نتیجه، پیچیدگی محاسباتی را کاهش می‌دهد. علاوه بر این، فرایند یادگیری را

1 Pooling layer.
2 Average
3 Max-pooling.
4 Mixed .
5 Importance-based pooling(LIP).
6 Stochastic Pooling(SP).
7 Soft function.
8 Discrete Wavelet Transform.
9 Objects Segmentation.
10 Training Phase.
11 Feed-forward.
12 Back-Propagation.
13 . Loss.
14 Threshold.
15 Stride.
16 Padding.

کنترل می‌کند و از بیش‌برازش، جلوگیری می‌کند. روش‌های مختلفی برای پیاده‌سازی لایه ادغام پیشنهاد شده است. غلامعلی-نژاد و همکاران برخی از آن‌ها را معرفی کرده‌اند [۳۰].

۲-۳. لایه اتصال کامل[1]

خروجی این لایه‌بردار، ویژگی ۱ بعدی است که با تبدیل ویژگی-های دو بعدی، مراحل قبل به‌دست می‌آید. این لایه‌ها، بیشترین پارامترها را در یک شبکه کانولوشن، مانند شبکه‌های عصبی مصنوعی سنتی تشکیل می‌دهد [۳۱]. این لایه اجازه می‌دهد تا

نتیجه شبکه در یک بردار با اندازه مشخص برای طبقه‌بندی یا پردازش بیشتر ارائه شود.

۳. مدل‌های لایه ادغام

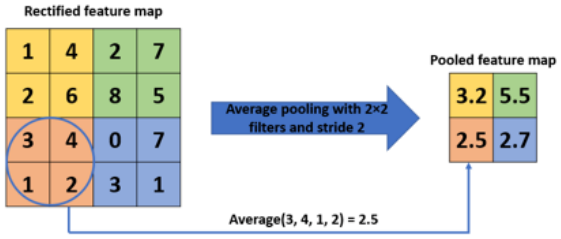
شکل ۱ نشان می‌دهد که لایه اول، ویژگی‌های ورودی را با استفاده از فیلترهای آموزش‌دیده، استخراج می‌کند. مجموعه‌ای از این لایه‌ها در یک مدل یادگیری عمیق برای سطح بالایی از یادگیری انتزاعی مانند اشیاء و اشکال خاص استفاده می‌شود [۳۲]. در نتیجه، حرکات کوچک مانند چرخش و برش در موقعیت ویژگی در تصویر ورودی منجر به نقشه‌های ویژگی مختلف می‌شود. نمونه‌گیری پایین[2] [۳۳] راه حل معروفی برای غلبه بر مشکل است. این روش سعی می‌کند ورودی‌هایی با کیفیت پایین‌تر، شامل عناصر ضروری را ایجاد کند. علاوه بر این، لایه کانولوشن سعی می‌کند با استفاده از تغییر کانولوشن روی تصویر به آن دست یابد.

یک روش قوی‌تر و رایج‌تر، استفاده از یک لایه ادغام است. به طور خاص، لایه کانولوشن یک تابع غیر خطی مانند ReLU را به خروجی اعمال می‌کند. این مدل، ترتیب لایه‌ها را در یک مدل شبکه عصبی کانولوشن نشان می‌دهد که می‌تواند یک یا چند بار تکرار شود. لایه ادغام یک نقشه ویژگی جدید را با پردازش هر نقشه ویژگی به طور جداگانه با تعداد یکسان تشکیل می‌دهد. ادغام مانند یک فیلتر برای ویژگی نقشه‌ها عمل می‌کند، به طوری که اندازه آن کوچک‌تر از اندازه نقشه ویژگی است. به عنوان مثال، ادغام با پیکسل‌های ۲×۲ را در نظر بگیرید که در آن تعداد گام ۲ پیکسل باشد. این ساختار، کاهش اندازه هر نقشه ویژگی را تا ۲ برابر نشان می‌دهد. علاوه بر این، ابعاد به نصف اندازه آن کاهش می‌یابد و تعداد پیکسل‌ها و مقادیر هر ویژگی یک چهارم، کاهش می‌یابد، بنابراین، اندازه نقشه ویژگی تلفیقی در خروجی ۳×۳ است وقتی که اندازه نقشه ویژگی ۶×۶ باشد. در نهایت، خلاصه‌ای از ویژگی‌های شناسایی‌شده در ورودی توسط یک لایه ادغام ایجاد می‌شود، و نقشه‌های ویژگی پایین نمونه‌برداری شده یا ترکیبی را ایجاد می‌کند. از آنجایی که ویژگی‌های کوچک در ورودی یک نقشه، ویژگی ترکیبی را با همان مکان ایجاد می‌کند، در لایه کانولوشن مفید است. در ادامه، توابع ادغام محبوب‌تر، به تفصیل ارائه شده است.

۱-۳. لایه ادغام میانگین

همان‌طور که از نام این تابع مشخص است، میانگین ویژگی‌ها را محاسبه می‌کند [۳۴]. به عنوان مثال، یک مقدار متوسط در هر

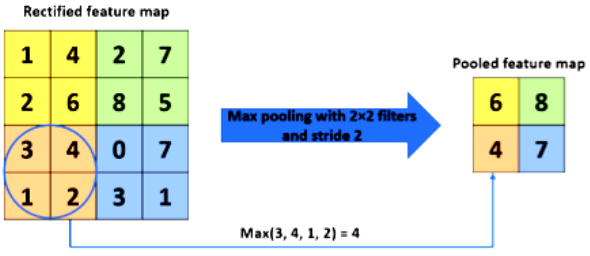
مربع برای نمونه‌برداری از مربع‌های ۲×۲ استفاده می‌شود. همان-طور که شکل ۳ نشان می‌دهد، روش نمونه‌برداری پایین در این روش ادغام، در دو مرحله انجام می‌شود: ابتدا ورودی به مناطق مستطیلی تقسیم می‌شود. سپس، مقدار متوسط برای هر منطقه محاسبه می‌شود و یک نقشه ویژگی جدید به عنوان خروجی ایجاد می‌شود.



شکل ۳: یک مثال از عملکرد لایه ادغام میانگین.

۲-۳. لایه ادغام ماکسیمم

این روش [۳۵] مشابه روش ادغام میانگین است، اما به جای مقدار میانگین، حداکثر مقدار را در هر منطقه پیدا می‌کند. شکل ۴ نمونه‌ای از این مدل را نشان می‌دهد.



شکل ۴: یک مثال از عملکرد لایه ادغام ماکسیمم.

1Fully Connected Layer.

2Down-sampling.

۳-۳. لایه ادغام ترکیبی

روش میانگین [۳۶] مقدار میانگین را بدون در نظر گرفتن حداکثر مقدار در خروجی محاسبه می‌کند. یو و همکاران، ساختار جدیدی به نام ادغام مختلط را پیشنهاد کرد که هم مدل متوسط و هم حداکثر را در برمی‌گیرد [۳۷]. معادله ۱ خروجی را محاسبه می‌کند.

$$s_j = \gamma \max_{i \in R_j} a_i + (1-\gamma)\frac{1}{|R_j|}\sum_{i \in R_j} a_i \qquad (1)$$

در معادله ۱ پارامتر γ نوع ادغام را تعیین می‌کند (γ =۰ برای ادغام متوسط و γ=۱ برای ادغام ماکسیمم). وزن ناحیه jام به صورت |Rj| نشان داده شده است. α_i عنصر i از ناحیه j ام است. مقدار γ برای ترتیب انتشار رو به جلو و در طی فرایند پس انتشار استفاده می‌شود.

۴-۳. لایه ادغام L_p

این روش، ابتدا نقشه ویژگی ورودی را به چند ناحیه مستطیلی جدا می‌کند. برای هر منطقه، میانگین وزنی عناصر منطقه را اختصاص می‌دهد. وزن ناحیه j تعیین کننده jامین ناحیه ادغام |Rj| است. Ai را به‌عنوان عنصر i امین ناحیه j در نظر بگیرید. بنابراین، معادله ۲ عنصر j خروجی را محاسبه می‌کند.

$$s_j = \left(\frac{1}{|R_j|}\sum_{i \in R_j} a_i^p\right)^{1/p} \qquad (2)$$

در معادله ۲، ρ یک مقدار بین ۱ و ∞ است. ρ پارامتری است که رفتار عملگر L_p را تعیین می‌کند. به عنوان مثال، اگر ρ=1 رفتار آن مشابه مدل ادغام متوسط است و ρ=∞ رفتاری مشابه با ادغام ماکسیمم دارد.

۵-۳. لایه ادغام تصادفی

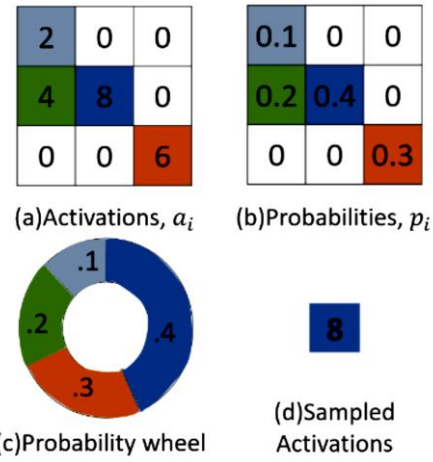
این مدل، بر اساس حذف تصادفی[1] و برای غلبه بر مشکل عدم اثربخشی حداکثر مقدار که مشکل اصلی ادغام میانگین است، طراحی شده است. این مدل، مقادیر تصادفی را با استفاده از توزیع چندجمله‌ای تولید می‌کند [۳۸]. این ادغام، ابتدا ورودی را به چند منطقه تقسیم می‌کند. سپس، احتمال ρ_i را برای هر منطقه بر اساس معادله ۳ محاسبه می‌کند.

$$p_i = \frac{a_i}{\sum_{k \in R_j} a_k} \qquad (3)$$

در معادله ۳، α_i عنصر i امین ناحیه jاست. این احتمالات یک توزیع چند جمله‌ای ایجاد می‌کند که l و α_l را مشخص می‌کند، به‌طوری‌که l مکان در منطقه jامین ادغام است و α_l عنصر مربوطه است. معادله ۴، Sj را که نقشه ویژگی خروجی عنصر jام است محاسبه می‌کند.

$$s_j = a_l \; where \; l \sim P(p_1, \dots, p_{|R_j|}) \qquad (4)$$

بنابراین، یک توزیع چند جمله‌ای، عناصر نقشه ویژگی خروجی را با استفاده از احتمالات محاسبه می‌کند. مؤلفه تصادفی، مدل ادغام تصادفی از بیش‌برازش جلوگیری می‌کند و شامل مزایای حداکثر ادغام می‌شود. شکل ۵ نمونه‌ای از رویه مدل ادغام تصادفی را نشان می‌دهد.



شکل ۵: یک مثال از عملکرد لایه ادغام تصادفی.

۶-۳. لایه ادغام بر اساس اهمیت محلّی

این مدل بر اساس مدل‌سازی اهمیت فضایی است. علاوه بر این، اهمیت محلّی، نقشه‌های با اهمیت پایین را در نمونه‌برداری کاهشی نامناسب در نظر می‌گیرد. این مدل، ویژگی‌ها را برای نمونه‌گیری کاهشی با یادگیری تمایزی[2] و نقشه‌های اهمیت تطبیقی جمع‌آوری می‌کند. این مدل براساس معادله ۵ محاسبه می شود:

$$LIP = \frac{\sum_{(\Delta x, \Delta y) \in \Omega} I_{x+\Delta x, y+\Delta y} \, exp(G(I))_{x+\Delta x, y+\Delta y}}{\sum_{(\Delta x, \Delta y) \in \Omega} exp(G(I))_{x+\Delta x, y+\Delta y}} \qquad (5)$$

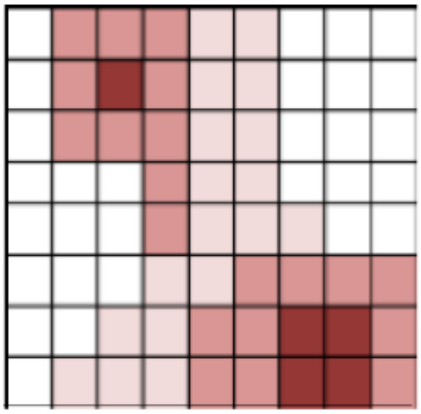
که در آن G(I) لگاریتم اهمیت است. I نقشه ویژگی ورودی را نشان می‌دهد. Ω مجموعه شاخص‌های هسته می‌باشد.

۷-۳. لایه ادغام نرم

1Dropout.

2discriminative learning.

این مدل، مجموعه‌ای از حداکثر عناصر فعال شده را در نقشه ویژگی [۴۳] در نظر می‌گیرد. شکل ۷ نمونه‌ای از ادغام نرم [۴۴] را نشان می‌دهد.



شکل ۶: یک مثال از عملکرد لایه ادغام نرم.

معادله ۶: خروجی ادغام نرم را در یک منطقه محاسبه می‌کند.

$$Soft(X) = \frac{ln[\frac{1}{|R_i|}\sum_{x_j \in R_i} exp(\lambda x_j)]}{\lambda} \qquad (6)$$

که در آن R_i منطقه i برای ورودی x است. λ یک فراپارامتر است. λ ادغام میانگین و ماکسیمم را محاسبه می‌کند به طوری‌که اگر λ->∞ آنگاه ادغام ماکسیمم و اگر λ->0 آنگاه ادغام میانگین را محاسبه می‌کند.

۳-۸. لایه ادغام DWT

این ساختار ترکیبی با استفاده از تبدیل موجک گسسته هار[1] [۴۵] فرموله شده است. این تبدیل به لایه ادغام اجازه می‌دهد تا ابعاد را کاهش دهد و ویژگی‌های مفید را استخراج کند [۴۶]. x را به عنوان یک تصویر در نظر بگیرید. یک ساختار دو بعدی DWT از چهار فیلتر، شامل سه بالا گذر (f_{HL}، f_{LH}، f_{HH}) و یک فیلتر پایین گذر (f_{LL}) برای ساخت چهار تصویر فرعی (x_{HH}، x_{LH}، x_{HL}، و x_{LL}) از تصویر استفاده می‌کند. معادلات ۷ و ۸ به طور دقیق فیلترها و تصاویر فرعی را تعریف می‌کند که هر کدام هدف خاصی را در فرایند تجزیه انجام می‌دهد.

$$\begin{cases} f_{LL} = \begin{bmatrix} +1 & +1 \\ +1 & +1 \end{bmatrix} & f_{LH} = \begin{bmatrix} -1 & -1 \\ +1 & +1 \end{bmatrix} \\ f_{HL} = \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix} & f_{HH} = \begin{bmatrix} +1 & -1 \\ -1 & +1 \end{bmatrix} \end{cases} \qquad (7)$$

$$\begin{cases} x_{LL} = (f_{LL} \otimes x) \downarrow_2 & x_{LH} = (f_{LH} \otimes x) \downarrow_2 \\ x_{HL} = (f_{HL} \otimes x) \downarrow_2 & x_{HH} = (f_{HH} \otimes x) \downarrow_2 \end{cases} \qquad (8)$$

که در آن ⊗ عملگر کانولوشن می‌باشد. $\downarrow_2$ یک عملگر استاندارد برای نمونه‌برداری پایین است و ضریب آن ۲ است. معادله ۹ مقدار (i,j) تصاویر خورشید را محاسبه می‌کند.

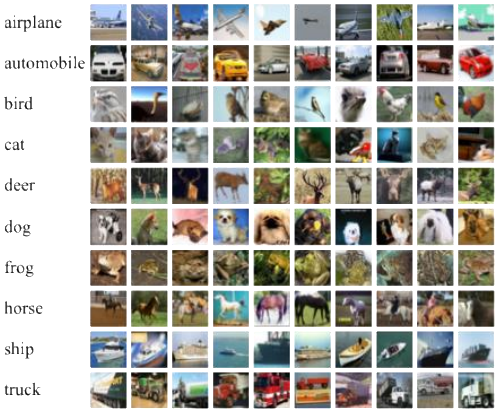
$$\begin{cases} x_{LL}(i,j) = x(2i-1,2j-1) + x(2i-1,2j) + \\ \qquad x(2i,2j-1) + x(2i,2j) \\ x_{LH}(i,j) = -x(2i-1,2j-1) - x(2i-1,2j) + \\ \qquad x(2i,2j-1) + x(2i,2j) \\ x_{HL}(i,j) = -x(2i-1,2j-1) + x(2i-1,2j) - \\ \qquad x(2i,2j-1) + x(2i,2j) \\ x_{HH}(i,j) = x(2i-1,2j-1) - x(2i-1,2j) - \\ \qquad x(2i,2j-1) + x(2i,2j) \end{cases} \qquad (9)$$

۴. ارزیابی

۴-۱. مجموعه داده

CIFAR-10 [47] و شبکه VGG16 برای ارزیابی مدل‌های مختلف ادغام استفاده شده‌اند. تعداد تصاویر رنگی ۶۰۰۰۰ ۳۲*۳۲ در ۱۰ دسته با ۶۰۰۰ تصویر است. تعداد تصاویر آموزشی ۵۰۰۰۰ و مجموعه تصاویر تست شامل ۱۰۰۰۰ می‌باشد. مجموعه داده، شامل پنج دسته به عنوان مجموعه آموزشی و یکی به عنوان آزمایش است. دسته‌بندی آزمون شامل ۱۰۰۰ تصویر است که به طور تصادفی از هر دسته انتخاب می‌شود. بقیه تصاویر، به صورت تصادفی دسته‌های آموزشی در نظر گرفته می‌شود، اگرچه تعداد تصاویر در دسته‌ها ممکن است متفاوت باشد. اما تعداد دسته‌های آموزشی دقیقاً ۵۰۰۰ عکس است که از هر دسته انتخاب می‌شود. شکل ۸ نمونه‌ای از مجموعه داده را نشان می‌دهد.

1Haar discrete wavelet transform.

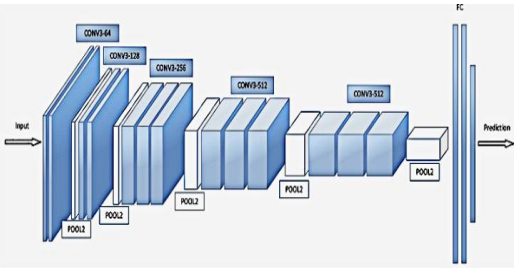


شکل ۷: نمونه‌ای از تصاویر CIFAR-10 [48].

به طور کلّی، یک تصویر شامل آرایه‌ای از مقادیر پیکسل است. این داده اختصاصی، حاوی اطلاعاتی در مورد تصاویر اصلی است. ویژگی-های آموزش دیده لایه اول نشان می‌دهد که آیا لبه‌هایی در جهت‌ها و مکان‌های خاص وجود دارد یا خیر؟ در لایه بعدی با وجود تغییرات جزئی در لبه، با شناسایی چینش‌های خاص لبه‌ها، نقوش را تشخیص می‌دهد. ویژگی‌های عمیق‌تر لایه سوم را می‌توان در ترکیب‌های بزرگ‌تری که مربوط به بخش‌هایی از اشیاء شناخته شده است، دسته‌بندی کرد. علاوه بر این، لایه‌های بعدی اشیاء مشابه را که ترکیبی از اشیاء شناخته شده ذکر شده است، تشخیص می‌دهد. ویژگی‌های آموزش‌دیده از لایه‌های مختلف مشتق شده‌اند که مفهومی مهم در روش‌های یادگیری عمیق است [۴۹].

۴-۲. ساختار شبکه

K. Simonyan و A. Zisserman یک شبکه عصبی کانولوشنال به نام شبکه VGG16 را پیشنهاد کردند [۵۰]. این شبکه توانست در رقابت معروف ImageNet به دقت بالایی دست یابد. معماری VGG 16 شامل ۱۶ لایه است که در شکل ۹ نشان داده شده است [۵۱]. علاوه بر این، شامل دو لایه کانولوشن است که شامل ۶۴ فیلتر ۳*۳ است. بعد از این لایه، یک لایه ادغام ماکسیمم در اندازه ۲*۲ و گام[1] ۲ اعمال می‌شود. لایه ادغام، اندازه ویژگی را به نصف کاهش می‌دهد. سپس، دو لایه کانولوشن وجود دارد که دارای ۱۲۸ فیلتر است. اندازه حداکثر ادغام این لایه‌ها ۳*۳ و ۲*۲ و گام ۲ می‌باشد. همچنین معماری، دارای سه لایه کانولوشن است که دارای ۲۵۶ فیلتر و اندازه آن ۳*۳ می‌باشد. بعد، یک ساختار ادغام با اندازه ۲ * ۲ و گام ۲ وجود دارد. شبکه با سه لایه کانولوشن که دارای ۵۱۲ فیلتر است ادامه می‌یابد و اندازه آن ۳ * ۳ است. سپس یک لایه ادغام ماکسیمم وجود دارد که دو بار تکرار می‌شود. در نهایت، یک بردار از ویژگی‌ها به لایه‌های کاملاً متصل ارائه می‌شود. دو لایه عصبی ۴۰۹۶ بعدی پشت سر هم قرار گرفته‌اند. در نهایت، یک لایه عصبی وجود دارد که شامل کلاس‌های کاربردی ما می‌شود. یک تابع فعال‌سازی به نام RELU در همه لایه‌ها اعمال می‌شود. شانزده در VGG 16 به معنای شانزده لایه بدون در نظر گرفتن لایه‌های ادغام می‌باشد. شکل ۹ معماری آن را نشان می‌دهد.



شکل ۸: معماری شبکه VGG-16.

۴-۳. محیط ارزیابی

کتابخانه Pytorch برای اجرای آزمایشات در پایتون استفاده شده است. سیستمی که برای اجرای بررسی‌ها انجام می‌شود، CPU core i3-9100f at 4000 operating frequency and RTX 2080 turbo graphics in gaming mode می‌باشد. آزمایش‌ها با استفاده از ادغام‌های مختلف در ساختار VGG16 انجام شد. در ادامه به توضیح معیارهای ارزیابی می‌پردازیم.

۴-۳-۱. صحت[2]

جدول ۱ توضیح اختصار در فرمول‌ها را نشان می‌دهد.

جدول ۱: پارامترهای مورد استفاده برای معرفی معیارهای طبقه‌بندی.

پارامتر	توصیف
TP[3]	تعداد داده‌های مثبت که به درستی مثبت شناسایی شده است.
FP[4]	تعداد داده‌های مثبتی که به اشتباه به عنوان منفی شناسایی شده‌اند.

1Step.
2Accuracy.

3True Positive: TP.
4False Positive: FP.

تعداد داده‌های منفی که به درستی منفی شناسایی شده‌اند.	TN[1]
تعداد داده‌های منفی که به اشتباه مثبت تشخیص داده شده‌اند.	FN[2]

صحت تشخیص، نسبت کلاس‌های پیش‌بینی‌شده درست به کل داده‌ها است. از نظر ریاضی می توان آن را به صورت زیر بیان کرد:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

۴-۳-۲. یادآوری[3]

زمانی‌که مقدار FN بالا باشد، معیار یادآوری مناسب خواهد بود. این معیار به صورت زیر تعریف می‌شود:

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

۴-۳-۳. دقت[4]

درستی یک نتیجه مثبت دقت در نظر گرفته می‌شود و حداکثر مقدار آن ۱ است. دقت بیشتر به معنای FP کمتر است. این معیار به صورت زیر تعریف می‌شود:

$$Precision = \frac{TP}{TP+FP} \quad (12)$$

۴-۳-۴. کاپا کوهن[5]

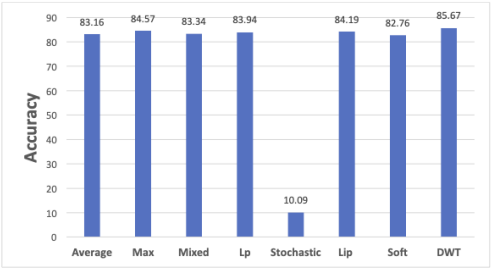
ضریب کاپا کوهن مقداری در [۱+، ۱-] است و هر چه به ۱+ نزدیک‌تر باشد، توافق[6] متناسب‌تر و مستقیم‌تر است. مقدار نزدیک ۱- نشان دهنده توافق معکوس است، در حالی که یک نزدیک به صفر، نشان-دهنده عدم توافق است. معادله ۱۳ ضریب [۵۲] را محاسبه می‌کند.

$$Kappa = \frac{Pr(a)-Pr(e)}{1-Pr(e)} \quad (13)$$

به‌طوری‌که Pr (a) قرارداد نسبی مشاهده شده بین ارزیاب‌ها است. Pr (e) به طور فرضی احتمال شانس قرارداد را محاسبه می‌کند.

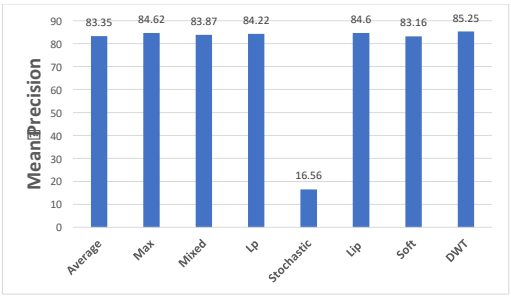
۵. نتایج و بحث

صحت نزدیکی یک مقدار اندازه‌گیری شده به یک مقدار معتبر است. به عبارت دیگر، نزدیک بودن یک اندازه‌گیری به مقدار استاندارد است. شکل ۱۰ نشان می‌دهد که VGG16 با BN بهترین نتایج را با استفاده ادغام ماکسیمم به‌دست آورد. علاوه بر این، استفاده از ادغام متوسط در طبقه‌بندی‌کننده آزمون، کمترین مقدار صحت را به همراه دارد.



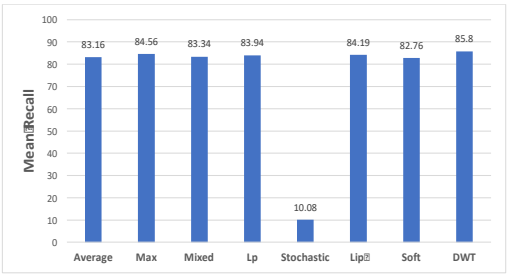
شکل ۹: مقایسه صحت مدل‌های ادغام.

دقت، نزدیکی اندازه‌گیری‌ها به یکدیگر است. خطای تصادفی و متغیرهای آماری را توصیف می‌کند. شکل ۱۱ دقت نتیجه آزمایشی را نشان می‌دهد. دقت، میزان نزدیکی اندازه‌گیری‌های مکرّر را در محیط یا شرایط یکسان نشان می‌دهد.



شکل ۱۰: مقایسه دقت مدل‌های ادغام.

یادآوری نقطه مقابل دقت است. بخشی از اسناد مرتبط را نشان می‌دهد که واقعاً بازیابی شده اند (شکل ۱۲ را ببینید).



شکل ۱۱: مقایسه میانگین فراخوانی مدل‌های ادغام.

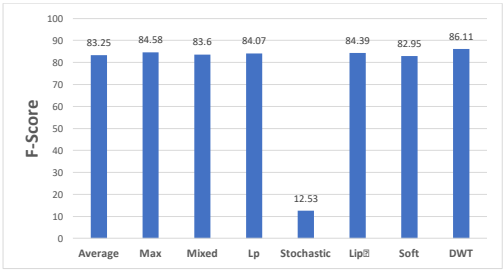
1True Negative: TN.
2False Negative: FN.
3Recall.
4Precision.

5Cohen Kappa.
6Agreement.

F-score هم دقت و هم یادآوری را برای یک کلاس مثبت خاص در نظر می‌گیرد. معادله ۱۴ امتیاز F را محاسبه می‌کند.

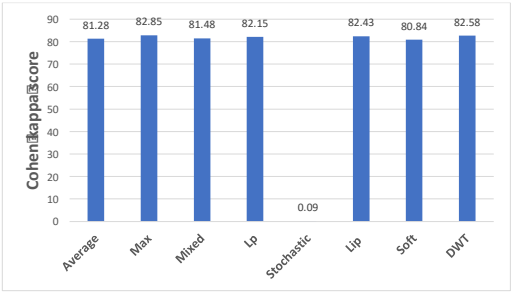
$$F - Score = 2 \times \frac{Precison \times Recall}{Precison + Recall} \qquad (14)$$

شکل ۱۳ مقایسه معیار F-score را نشان می‌دهد. در این آزمون، ادغام ماکسیمم حداکثر مقدار را تولید کرده است. همچنین ادغام میانگین کمترین مقدار را داشته است.



شکل ۱۲: مقایسه معیار F-score مدل‌های ادغام.

کاپا کوهن، درجه انطباق اندازه‌گیری‌ها در شرایط مختلف برای متغیرهایی است که یکسان هستند. این اندازه‌گیری به ما می‌گوید که طبقه‌بندی کننده آزمایش شده در مقایسه با طبقه‌بندی کننده‌ای که به طور تصادفی با توجه به فراوانی هر کلاس حدس می‌زند، چقدر بهتر است. دقت یک معیار شهودی است که برای بررسی عملکرد طبقه‌بندی استفاده می‌شود، اما فقط با مجموعه داده‌های متعادل کار می‌کند. از سوی دیگر، کاپا کوهن، این نوع مجموعه داده را مدیریت می‌کند. می‌توان گفت که این یک اندازه‌گیری، نرمال شده از دقت است و به مدل اجازه می‌دهد تا کلاس اقلیت را حتی زمانی که احتمال پیش‌بینی شده آن کلاس بالاترین نیست، پیش‌بینی کند. شکل ۱۴ نتیجه را بر اساس معیار کاپا کوهن نشان می‌دهد.



شکل ۱۳: مقایسه معیارکاپا کوهن مدل‌های

این مقاله از شبکه عصبی VGG16 برای اندازه‌گیری عملکرد روش‌ها استفاده کرده است. به دلیل ساختار عمیق و تعداد بالای عناصر قابل آموزش، آموزش آن تنها با نرمال‌سازی دسته‌ای امکان-پذیر است. همان‌طور که در جدول ۲ نشان داده شده است، لایه-های ادغام مختلف، زمان‌های شناسایی مختلفی را برای VGG16 ایجاد کردند. کندترین حالت زمان تصادفی است که ۰/۷۹۵ میلی ثانیه است. با وجود زمان تشخیص طولانی، این ادغام هنوز به یادگیری مطلوب نرسیده است. سریع‌ترین تشخیص با استفاده از روش ادغام ماکسیمم به دست آمده است. این روش از نظر معیارهای طبقه‌بندی بهترین بود. پس از جمع‌آوری حداکثر، روش ادغام Lip بهترین نتایج را دارد، اما زمان تشخیص را تقریباً پنجاه درصد افزایش داده است. این افزایش زمان به دلیل پیچیدگی‌های محاسباتی لگاریتمی و نمایی آن است. روش ترکیبی در آزمایش‌ها نمی‌توانست بهتر از ادغام ماکسیمم، در مورد معیارهای طبقه‌بندی رفتار کند. اما این روش، کمترین مقدار خطا[1]را در مرحله اموزش دارد. همان‌طور که مشاهده می‌شود، سریع‌ترین و بهترین روش، روش ادغام ماکسیمم است که از نظر محاسباتی ساده‌ترین روش است.

جدول ۲: نتایج عملکرد لایه‌های مختلف ادغام.

Pooling type	Train loss	Test accuracy	Test loss	F-score	Mean precision	Mean recall	Cohen kappa score	
Average	0/5924	83/16	0/6309	83/25	83/35	83/16	81/28	0/296
Max	0/5168	84/57	0/5930	84/58	84/62	84/56	82/85	0/291
Mixed	0/5043	83/34	0/6654	83/60	83/87	83/34	81/48	0/318
Lp	0/5307	83/94	0/6047	84/07	84/22	83/94	82/15	0/314
Stochastic	1/8256	10/09	2/3088	12/53	16/56	10/08	0/09	0/795
Lip	0/5329	84/19	0/6116	84/39	84/60	84/19	82/43	0/435
Soft	0/5686	82/76	0/6987	82/95	83/16	82/76	80/84	0/331
DWT	0/2685	85/67	0/3568	86/11	85/25	85/80	82/58	0/524

1.L oss

۶. نتیجه‌گیری

این مقاله، به بررسی روش‌های مختلف ادغام محبوب برای مسائل و کاربردهای طبقه‌بندی پرداخته است. سپس، توانایی روش‌های ادغام را در ساختار VGG16 در مجموعه داده CIFAR-10 بررسی کرده است. این مطالعه شبکه VGG16را برای پارامترهای معیارهای طبقه-بندی شامل صحت، دقت، یادآوری، F1 -score و امتیاز کوهن کاپا آزمایش می‌کند. روش‌های ادغام، ابعاد نقشه‌های ویژگی و استخراج ویژگی هدف را کاهش می‌دهد. دلیل تفاوت در معیارهای طبقه‌بندی، استخراج هدفمند ویژگی است. نتایج نشان داد که بهترین لایه، ادغام ماکسیمم برای کاربردهای طبقه‌بندی می‌باشد. کارهای آینده نقش روش‌های ادغام در کاربردهای مختلف را بررسی خواهند کرد.

منابع

[1]. Z. Lu et al., "Multiobjective Evolutionary Design of Deep Convolutional Neural Networks for Image Classification", IEEE Transactions on Evolutionary Computation, vol. 25, no. 2, pp. 277-291, 2020.

[2]. H. Gholamalinejad and H. Khosravi, "IRVD: A Large-Scale Dataset for Classification of Iranian Vehicles in Urban Streets", Journal of AI and Data Mining, vol. 9, no. 1, pp. 1-9, 2021.

[3]. A. Pal et al., "Deep Metric Learning for Cervical Image Classification", IEEE Access, vol. 9, pp. 53266-53275, 2021.

[4]. Q. Gao, T. Long, and Z. Zhou, "Mineral identification based on natural feature-oriented image processing and multi-label image classification", Expert Systems with Applications, vol. 238, p. 122111, 2024.

[5]. P. Wang, E. Fan, and P. Wang, "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning", Pattern Recognition Letters, vol. 141, pp. 61-67, 2021.

[6]. T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li, "Bag of tricks for image classification with convolutional neural networks", in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 558-567, 2019.

[7]. H. A. Alsattar et al, "Developing deep transfer and machine learning models of chest X-ray for diagnosing COVID-19 cases using probabilistic single-valued neutrosophic hesitant fuzzy", Expert Systems with Applications, vol. 236, p. 121300, 2024.

[8]. A. Goel, A. K. Goel, and A. Kumar, "The role of artificial neural network and machine learning in utilizing spatial information", Spatial Information Research, vol. 31, no. 3, pp. 275-285, 2023.

[9]. N. Alturayeif, H. Luqman, and M. Ahmed, "A systematic review of machine learning techniques for stance detection and its applications", Neural Computing and Applications, vol. 35, no. 7, pp. 5113-5144, 2023.

[10]. W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, "Definitions, methods, and applications in interpretable machine learning", Proceedings of the National Academy of Sciences, vol. 116, no. 44, pp. 22071-22080, 2019.

[11]. C. F. F. Sousa, "Categories' churn: a machine learning approach in retail", 2023.

[12]. M. Iyer and R. Mehra, "Forecasting Price of Indian Stock Market Using Supervised Machine Learning Technique", in Progress in Advanced Computing and Intelligent Engineering: Springer, pp. 3-17, 2021.

[13]. K. A. Taher, B. M. Y. Jisan, and M. M. Rahman, "Network intrusion detection using supervised machine learning technique with feature selection", in 2019 International conference on robotics, electrical and signal processing techniques (ICREST), IEEE, pp. 643-646, 2019.

[14]. L. Rundo et al, "Automated prostate gland segmentation based on an unsupervised fuzzy C-means clustering technique using multispectral T1w and T2w MR imaging", Information, vol. 8, no. 2, p. 49, 2017.

[15]. J. H. Park and S. I. Cho, "The Analysis of Semi-supervised Learning Technique of Deep Learning-based Classification Model", Journal of Broadcast Engineering, vol. 26, no. 1, pp. 79-87, 2021.

[16]. V. Cheplygina, M. de Bruijne, and J. P. Pluim, "Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis", Medical image analysis, vol. 54, pp. 280-296, 2019.

[17]. D. Kangin and N. Pugeault, "Combination of Supervised and Reinforcement Learning For Vision-Based Autonomous Control", 2018.

[18]. J. Wang, T. Yang, L. Tang, Y. Wang, and R. Xiong, "Learning an Efficient and Safe Policy for Highway Driving using Supervised Learning and Reinforcement Learning", in 2019 IEEE International Conference on Real-time Computing and Robotics (RCAR), IEEE, pp. 112-117, 2019.

[19]. N. O. Nikitin, I. S. Polonskaia, P. Vychuzhanin, I. V. Barabanova, and A. V. Kalyuzhnaya, "Structural evolutionary learning for composite classification models", Procedia Computer Science, vol. 178, pp. 414-423, 2020.

[20]. A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection", Progress in Artificial Intelligence, vol. 9, no. 2, pp. 85-112, 2020.

[21]. N. Ma and P. Chawlab, "Deep CNN architectures for learning image classification: A systematic review, taxonomy and open challenges", in Artificial Intelligence and Speech Technology: Proceedings of the 2nd International Conference on Artificial Intelligence and

Classification," in International Conference on Intelligent Systems Design and Applications, pp. 171-180, 2020.

[35]. V. Christlein, L. Spranger, M. Seuret, A. Nicolaou, P. Král, and A. Maier, "Deep generalized max pooling", in 2019 International Conference on Document Analysis and Recognition (ICDAR), pp. 1090-1096, 2019.

[36]. T. Sharma, N. K. Verma, and S. Masood, "Mixed fuzzy pooling in convolutional neural networks for image classification", Multimedia Tools and Applications, vol. 82, no. 6, pp. 8405-8421, 2023.

[37]. D. Yu, H. Wang, P. Chen, and Z. Wei, "Mixed pooling for convolutional neural networks", in International conference on rough sets and knowledge technology, pp. 364-375, 2014.

[38]. U. Nandi, A. Ghorai, M. M. Singh, C. Changdar, S. Bhakta, and R. Kumar Pal, "Indian sign language alphabet recognition system using CNN with diffGrad optimizer and stochastic pooling", Multimedia Tools and Applications, vol. 82, no. 7, pp. 9627-9648, 2023.

[39]. P. Msonda, S. A. Uymaz, and S. S. Karaağaç, "Spatial Pyramid Pooling in Deep Convolutional Networks for Automatic Tuberculosis Diagnosis", Traitement du Signal, vol. 37, no. 6, 2020.

[40]. C. Dewi, R.-C. Chen, H. Yu, and X. Jiang, "Robust detection method for improving small traffic sign recognition based on spatial pyramid pooling", Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 7, pp. 8135-8152, 2023.

[41]. K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition", IEEE transactions on pattern analysis and machine intelligence, vol. 37, no. 9, pp. 1904-1916, 2015.

[42]. J. Zhou, Y. Tian, C. Yuan, K. Yin, G. Yang, and M. Wen, "Improved uav opium poppy detection using an updated yolov3 model", Sensors, vol. 19, no. 22, p. 4851, 2019.

[43]. I. Hwang and K. Lee, "Removing Speaker Information from Speech Representation using Variable-Length Soft Pooling", arXiv preprint arXiv:2404.00856, 2024.

[44]. A. Stergiou, R. Poppe, and G. Kalliatakis, "Refining activation downsampling with Softpool", arXiv preprint arXiv:2101.00440, 2021.

[45]. H. Gholamalinejad and H. Khosravi, "Vehicle classification using a real-time convolutional structure based on DWT pooling layer and SE blocks", Expert Systems with Applications, vol. 183, p. 115420, 2021.

[46]. X. Ma, D. Wang, D. Liu, and J. Yang, "DWT and CNN based multi-class motor imagery electroencephalographic signal recognition", Journal of neural engineering, vol. 17, no. 1, p. 016073, 2020.

[47]. W. Cukierski, "CIFAR-10-Object Recognition in Images", Kaggle 2013Available online: https://www/. Speech Technology,(AIST2020), 19-20 November, 2020, Delhi, India, CRC Press, p. 121, 2021.

[22]. P. P. Liang, A. Zadeh, and L.-P. Morency, "Foundations & trends in multimodal machine learning: Principles, challenges, and open questions", ACM Computing Surveys, vol. 56, no. 10, pp. 1-42, 2024.

[23]. A. Ardakani, C. Condo, M. Ahmadi, and W. J. Gross, "An architecture to accelerate convolution in deep neural networks", IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 65, no. 4, pp. 1349-1362, 2017.

[24]. H. Gholamalinejad and H. Khosravi, "Vehicle Classification using a Real-Time Convolutional Structure based on DWT pooling layer and SE blocks", Expert Systems with Applications, p. 115420, 2021.

[25]. C. Dias et al, "Simulating the behaviour of choquet-like (pre) aggregation functions for image resizing in the pooling layer of deep learning networks", in International Fuzzy Systems Association World Congress, pp. 224-236, 2019.

[26]. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs", IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 4, pp. 834-848, 2017.

[27]. J. Jordan, "Common architectures in convolutional neural networks", Режим доступу: https://www/. jeremyjordan. me/convnet-architectures, 2018.

[28]. Z. He, "Deep Learning in Image Classification: A Survey Report", in 2020 2nd International Conference on Information Technology and Computer Application (ITCA), , pp. 174-177, 2020.

[29]. C. Luo, S. Su, Y. Sun, Q. Tan, M. Han, and Z. Tian, "A convolution-based system for malicious URLs detection", Comput, Mater. Continua, vol. 62, no. 1, pp. 399-411, 2020.

[30]. H. Gholamalinezhad and H. Khosravi, "Pooling methods in deep neural networks, a review", arXiv preprint arXiv:2009.07485, 2020.

[31]. Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review", Neurocomputing, vol. 187, pp. 27-48, 2016.

[32]. M. Imani and H. Ghassemian, "An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges". Information fusion, vol. 59, pp. 59-83, 2020.

[33]. M.-M. Ho, G. He, Z. Wang, and J. Zhou, "Down-sampling based video coding with degradation-aware restoration-reconstruction deep neural network". in International Conference on Multimedia Modeling, pp. 99-110, 2020.

[34]. A. Al-Sabaawi, H. M. Ibrahim, Z. M. Arkah, M. Al-Amidie, and L. Alzubaidi, "Amended Convolutional Neural Network with Global Average Pooling for Image".

kaggle. com/c/cifar-10 (accessed on 30 December 2022), 2013.

[48]. A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images", 2009.

[49]. S. Hinterstoisser, V. Lepetit, P. Wohlhart, and K. Konolige, "On pre-trained image features and synthetic images for deep learning", in Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018.

[50]. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556, 2014.

[51]. T. Sulistyowati, P. PURWANTO, F. Alzami, and R. A. Pramunendar, "Vgg16 deep learning architecture using imbalance data methods for the detection of apple leaf diseases", Moneter: Jurnal Keuangan dan Perbankan, vol. 11, no. 1, pp. 41-53, 2023.

[52]. S. M. Vieira, U. Kaymak, and J. M. Sousa, "Cohen's kappa coefficient as a performance measure for feature selection", in International conference on fuzzy systems, pp. 1-8, 2010.

Exploring the types of fusion layers for deep learning-based image classification

Azam Noroozi[1*]

1- Faculty of Engineering - Department of Civil Engineering and Architecture - Torbath Heydariyeh University – Torbath Heydariyeh - Iran.

Abstract

The fusion layer has been widely used for image classification with more than a decade of deep learning. Many models using the fusion layer have tried to overcome the disadvantages of the traditional algorithm, which depends on the hand-drawn features. In general, deep learning models often contain trainable parameters and require a large number of labeled examples to achieve optimal performance. This paper focuses on optimal parameters and related previous work. Various advanced approaches have been evaluated through several experiments. The evaluation results on the CIFAR-10 dataset showed that the fastest and best method is the computationally simplest method, which is the Maximum Fusion method. After the Maximum method, the Lip Fusion method has the best results, but it increases the detection time by almost fifty percent. This increase in time is due to its logarithmic and exponential computational complexities. The hybrid method could not perform better than the maximum pooling in terms of classification criteria in experiments. However, this method has the lowest error rate in the training phase.

Keywords: Deep learning, fusion layer, image classification, convolutional neural network, machine learning.