

## بررسی انواع لایه‌های ادغام برای طبقه‌بندی تصاویر مبتنی بر یادگیری عمیق

اعظم نوروزی<sup>\*</sup>

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

### اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۳/۱۰/۲۱

تاریخ پذیرش: ۱۴۰۳/۱۲/۰۶

واژه‌های کلیدی: یادگیری عمیق، لایه ادغام، طبقه‌بندی تصویر، شبکه عصبی کانولوشن، یادگیری ماشین.

### 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 کلی، مدل‌های یادگیری عمیق، اغلب حاوی پارامترهای قابل آموزش هستند و برای دستیابی به عملکرد بهینه به نمونه‌های برچسب‌گذاری شده 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 CIFAR-10 نشان داد که سریع‌ترین و بهترین روش، ساده‌ترین روش از نظر محاسباتی می‌باشد که روش ادغام ماکسیمم است. پس از روش ماکسیمم، 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 را در مرحله آموزش دارد.

<sup>\*</sup> نویسنده مسئول، پست الکترونیکی: [a.noroozi@torbath.ac.ir](mailto:a.noroozi@torbath.ac.ir)

## ۱. مقدمه

## ۱.۱. پیشینه

طبقه‌بندی تصویر، یک فرایند پیچیده است که عوامل زیادی را در نظر می‌گیرد [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 کلی، تکنیک‌های یادگیری ماشین به صورت زیر طبقه‌بندی می‌شوند [۱۰، ۱۱]: روش‌های یادگیری تحت نظارت<sup>۲</sup> [۱۲، ۱۳] که مجموعه داده‌ها را بر اساس مدل‌ها آموزش می‌دهد. انواع اصلی این روش‌ها طبقه‌بندی و رگرسیون<sup>۳</sup> است. الگوریتم‌های یادگیری بدون نظارت<sup>۴</sup> [۱۴]، سعی می‌کنند شباهت بین داده‌های ورودی را بیابند. خوشه‌بندی<sup>۵</sup>، روش مناسب برای این دسته است. یادگیری نیمه نظارت شده<sup>۶</sup> [۱۵، ۱۶] از داده‌های برچسب‌دار و بدون برچسب استفاده می‌کند. این تکنیک، یک نوع از مدل یادگیری نظارت شده است. یادگیری تقویتی تحت نظارت<sup>۷</sup> [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<sup>۸</sup> به دلیل دقت بالا، یک مدل پیشرو برای خوشه‌بندی، تشخیص و طبقه‌بندی تصاویر است [۶، ۲۰]. روش‌های یادگیری عمیق با ترکیب بخش‌های<sup>۹</sup> ساده و غیر خطی، نمایش یک سطحی را به سطح بالاتر تبدیل می‌کند. در نتیجه، این تبدیل‌های 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 کمک‌کننده.
- تعیین تعداد نورون‌ها در تمام لایه‌های پنهان.
- تعیین تعداد لایه‌های پنهان<sup>۱۰</sup>.
- توابع فعال‌سازی<sup>۱۱</sup>.
- روش‌های بهینه‌سازی.
- تابع محوسازی<sup>۱۲</sup>.
- تعیین تعداد دوره‌ها<sup>۱۳</sup> و اندازه دسته<sup>۱۴</sup> 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، توجه تحقیقاتی گسترده‌ای را به خود جلب کرده است، بنابراین، مدل‌های مختلفی را برای کاربردهای مختلف پیشنهاد کرده‌اند [۲۳]. مدل شبکه عصبی کانولوشن، ابعاد نقشه<sup>۱۵</sup> ویژگی را کاهش می‌دهد و در نتیجه، پیچیدگی زمانی کمتری دارد [۲۴، ۲۵].

## ۱.۳. اهداف

این مقاله، یک بررسی کامل و جامع از انواع لایه‌های پرکاربرد ادغام<sup>۱۶</sup> پیشنهاد شده در روش‌های یادگیری عمیق در طبقه‌بندی تصاویر را انجام می‌دهد. ابتدا محبوب‌ترین لایه‌های ادغام شامل میانگین<sup>۱۷</sup>

<sup>۹</sup>. Module.

<sup>۱۰</sup>. Hidden Layer.

<sup>۱۱</sup>. Activation Function.

<sup>۱۲</sup>. Decay Function.

<sup>۱۳</sup>. Epoch.

<sup>۱۴</sup>. Batch size.

<sup>۱۵</sup>. Feature map.

<sup>۱۶</sup>. Pooling layer.

<sup>۱۷</sup>. Average

<sup>۱</sup>. Classification.

<sup>۲</sup>. Supervised Learning.

<sup>۳</sup>. Regression.

<sup>۴</sup>. Unsupervised Learning.

<sup>۵</sup>. Clustering.

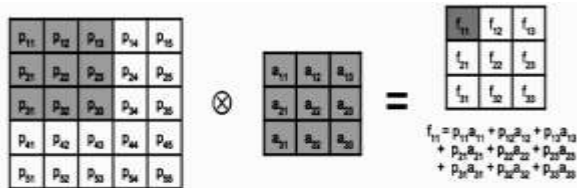
<sup>۶</sup>. Semi-supervised Learning.

<sup>۷</sup>. Reinforcement Learning.

<sup>۸</sup>. Convolution Neural Network.

## 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 توجه به اینکه این لایه با استفاده از روش‌هایی مانند گام حرکت<sup>۱۳</sup> یا لبه-گذاری<sup>۱۴</sup> 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 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].

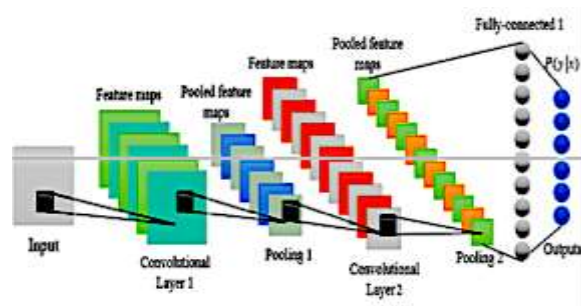
## ۲.۳. لایه اتصال کامل<sup>۱۵</sup>

خروجی این لایه‌بردار، ویژگی<sup>۱</sup> 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

ماکسیم<sup>۱</sup>، ترکیبی<sup>۲</sup>، Lp، ادغام مبتنی بر اهمیت محلی<sup>۳</sup>، تصادفی<sup>۴</sup>، تابع نرم<sup>۵</sup> و DWT<sup>۶</sup> 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 مدل-های 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 طبقه-بندی و بخش‌بندی اشیاء<sup>۷</sup> 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<sup>۸</sup> در این شبکه‌ها، مرحله پیش انتشار<sup>۹</sup> 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 لایه‌ها 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<sup>۱۰</sup> آغاز می‌شود. مرحله آموزش تا زمانی که میزان خطا<sup>۱۱</sup> به یک آستانه<sup>۱۲</sup> 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.

<sup>۹</sup>. Feed-forward.

<sup>۱۰</sup>. Back-Propagation.

<sup>۱۱</sup>. Loss.

<sup>۱۲</sup>. Threshold.

<sup>۱۳</sup>. Stride.

<sup>۱۴</sup>. Padding.

<sup>۱۵</sup>. Fully Connected Layer.

<sup>۱</sup>. Max-pooling.

<sup>۲</sup>. Mixed.

<sup>۳</sup>. Importance-based pooling (LIP).

<sup>۴</sup>. Stochastic Pooling (SP).

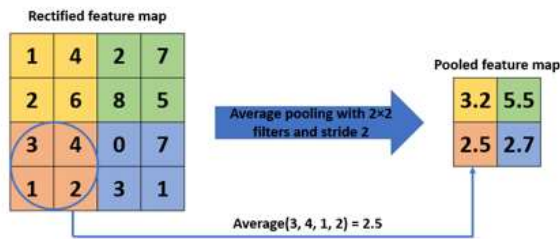
<sup>۵</sup>. Soft function.

<sup>۶</sup>. Discrete Wavelet Transform.

<sup>۷</sup>. Objects Segmentation.

<sup>۸</sup>. Training Phase.

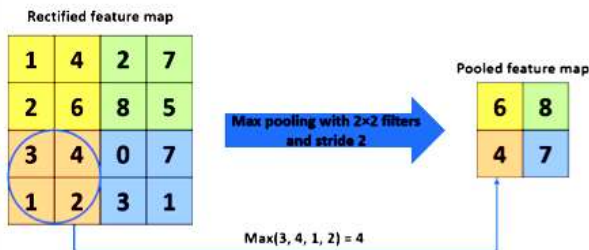
همان‌طور که از نام این تابع مشخص است، میانگین ویژگی‌ها را محاسبه می‌کند [۳۴]. به عنوان مثال، یک مقدار متوسط در هر مربع برای نمونه‌برداری از مربع‌های  $2 \times 2$  استفاده می‌شود. همان‌طور که شکل ۳ نشان می‌دهد، روش نمونه‌برداری پایین در این روش ادغام، در دو مرحله انجام می‌شود: ابتدا ورودی به مناطق مستطیلی 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.

### ۳.۲. 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.

### 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.

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

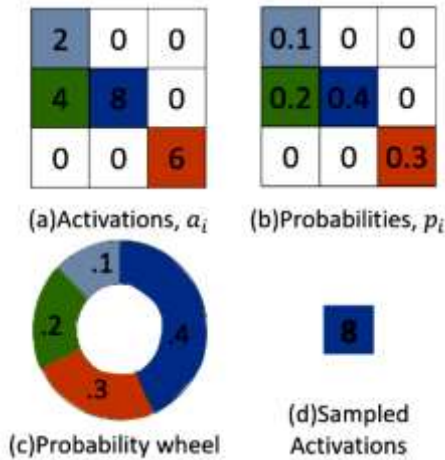
شکل ۱ نشان می‌دهد که لایه اول، ویژگی‌های ورودی را با استفاده از فیلترهای آموزش‌دیده، استخراج می‌کند. مجموعه‌ای از این لایه‌ها در یک مدل یادگیری عمیق برای سطح بالایی از یادگیری انتزاعی مانند اشیاء و اشکال خاص استفاده می‌شود [۳۲]. در نتیجه، حرکات کوچک مانند چرخش و برش در موقعیت ویژگی در تصویر ورودی منجر به نقشه‌های ویژگی مختلف می‌شود. نمونه‌گیری پایین<sup>۱</sup> [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  $2 \times 2$  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  $3 \times 3$  است وقتی که اندازه نقشه ویژگی  $6 \times 6$  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.

### ۳.۱. لایه ادغام میانگین

۱. Down-sampling.

ادغام می‌شود. شکل ۵ نمونه‌ای از رویه مدل ادغام تصادفی را نشان می‌دهد.



شکل ۵. یک مثال از عملکرد لایه ادغام تصادفی.

### ۳.۶. لایه ادغام بر اساس اهمیت محلی

این مدل بر اساس مدل‌سازی اهمیت فضایی است. علاوه بر این، اهمیت محلی، نقشه‌های با اهمیت پایین را در نمونه‌برداری کاهشی نامناسب در نظر می‌گیرد. این مدل، ویژگی‌ها را برای نمونه‌گیری کاهشی با یادگیری تمایزی<sup>۲</sup> و نقشه‌های اهمیت تطبیقی جمع‌آوری می‌کند. این مدل براساس معادله ۵ محاسبه می‌شود:

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

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

### ۳.۷. لایه ادغام نرم

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.

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

در معادله ۱ پارامتر  $\gamma$  نوع ادغام را تعیین می‌کند ( $\gamma = 0$  برای ادغام متوسط و  $\gamma = 1$  برای ادغام ماکسیمم). وزن ناحیه  $z$ ام به صورت  $|R_j|$  نشان داده شده است.  $a_i$  عنصر  $i$  از ناحیه  $z$ ام است. مقدار  $\gamma$  برای ترتیب انتشار رو به جلو و در طی فرایند پس انتشار استفاده می‌شود.

### ۳.۴. 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  $z$  تعیین کننده زامین ناحیه ادغام  $|R_j|$  است.  $A_i$  را به عنوان عنصر  $i$  امین ناحیه  $z$  در نظر بگیرید. بنابراین، معادله ۲ عنصر  $z$  خروجی را محاسبه می‌کند.

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

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

### ۳.۵. لایه ادغام تصادفی

این مدل، بر اساس حذف تصادفی<sup>۱</sup> 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  $p_i$  را برای هر منطقه بر اساس معادله ۳ محاسبه می‌کند.

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

در معادله ۳،  $a_i$  عنصر  $i$  امین ناحیه  $z$  است. این احتمالات یک توزیع چند جمله‌ای ایجاد می‌کند که  $l$  و  $a_1$  را مشخص می‌کند، به طوری که  $l$  مکان در منطقه زامین ادغام است و  $a_1$  عنصر مربوطه است. معادله ۴،  $s_j$  را که نقشه ویژگی خروجی عنصر  $z$ ام است محاسبه می‌کند.

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

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

<sup>۲</sup>. discriminative learning.

<sup>۱</sup>. Dropout.

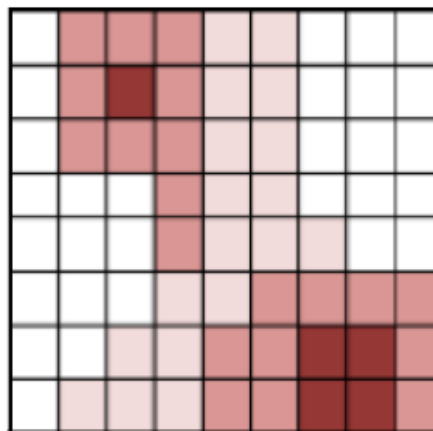
که در آن  $\otimes$  عملگر کانولوشن می‌باشد.  $\downarrow_2$  یک عملگر استاندارد برای نمونه‌برداری پایین است و ضرب آن ۲ است. معادله ۹ مقدار  $(i,j)$  تصاویر خورشید را محاسبه می‌کند.

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

#### ۴. ارزیابی

##### ۴.۱. مجموعه داده

[۴۷] CIFAR-۱۰ و شبکه VGG<sup>۱۶</sup> برای ارزیابی مدل‌های مختلف ادغام استفاده شده‌اند. تعداد تصاویر رنگی ۶۰۰۰۰  $32 \times 32$  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.



شکل ۶: یک مثال از عملکرد لایه ادغام نرم.

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

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

که در آن  $R_i$  منطقه  $i$  برای ورودی  $x$  است.  $\lambda$  یک فرایارامتر است.  $\lambda$  ادغام میانگین و ماکسیمم را محاسبه می‌کند به طوری که اگر  $\lambda \rightarrow \infty$  آنگاه ادغام ماکسیمم و اگر  $\lambda \rightarrow 0$  آنگاه ادغام میانگین را محاسبه می‌کند.

##### ۳.۸. لایه ادغام DWT

این ساختار ترکیبی با استفاده از تبدیل موجک گسسته هار<sup>۱</sup> [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 ( $f_{HH}, f_{LH}, f_{HL}$ ) و یک فیلتر پایین گذر ( $f_{LL}$ ) برای ساخت چهار تصویر فرعی ( $x_{LL}, x_{LH}, x_{HL}, x_{HH}$ ) 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.

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

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

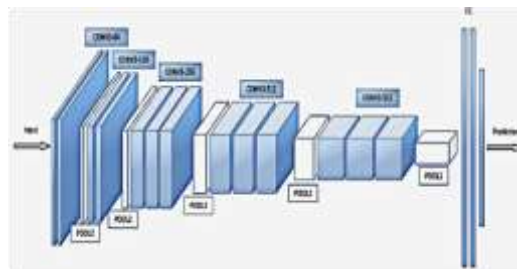
<sup>۱</sup>. Haar discrete wavelet transform.



کانولوشن که دارای ۵۱۲ فیلتر است ادامه می‌یابد و اندازه آن  $3 \times 3$  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 ۱۶ به معنای شانزده لایه بدون در نظر گرفتن لایه‌های ادغام می‌باشد. شکل ۹ معماری آن را نشان می‌دهد.



شکل ۷: نمونه‌ای از تصاویر CIFAR-۱۰ [۴۸].



شکل ۸: معماری شبکه VGG-۱۶.

### 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 turbo graphics in gaming mode 2080 می‌باشد. آزمایش‌ها با استفاده از ادغام‌های مختلف در ساختار VGG-16 انجام شد. در ادامه به توضیح معیارهای ارزیابی می‌پردازیم.

#### ۴.۳.۱. صحت<sup>۲</sup>

جدول ۱ توضیح اختصار در فرمول‌ها را نشان می‌دهد.

جدول ۱: پارامترهای مورد استفاده برای معرفی معیارهای طبقه‌بندی.

| توصیف                                                 | پارامتر         |
|-------------------------------------------------------|-----------------|
| تعداد داده‌های مثبت که به درستی مثبت شناسایی شده است. | TP <sup>۳</sup> |

۳. True Positive: TP.

به طور کلی، یک تصویر شامل آرایه‌ای از مقادیر پیکسل است. این داده اختصاصی، حاوی اطلاعاتی در مورد تصاویر اصلی است. ویژگی‌های آموزش دیده لایه اول نشان می‌دهد که آیا لایه‌هایی در جهت‌ها و مکان‌های خاص وجود دارد یا خیر؟ در لایه بعدی با وجود تغییرات جزئی در لایه، با شناسایی چینش‌های خاص لایه‌ها، 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].

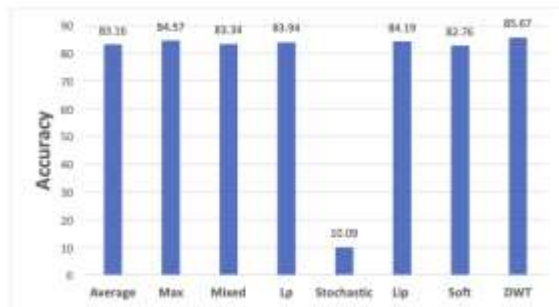
### ۴.۲. ساختار شبکه

K. Simonyan و A. Zisserman یک شبکه عصبی کانولوشنال به نام شبکه VGG-16 را پیشنهاد کردند [۵۰]. این شبکه توانست در رقابت معروف ImageNet به دقت بالایی دست یابد. معماری VGG ۱۶ 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  $3 \times 3$  است. بعد از این لایه، یک لایه ادغام ماکسیمم در اندازه  $2 \times 2$  و گام 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  $3 \times 3$  و  $2 \times 2$  و گام ۲ می‌باشد. همچنین معماری، دارای سه لایه کانولوشن است که دارای ۲۵۶ فیلتر و اندازه آن  $3 \times 3$  می‌باشد. بعد، یک ساختار ادغام با اندازه  $2 \times 2$  و گام ۲ وجود دارد. شبکه با سه لایه

۱. Step.

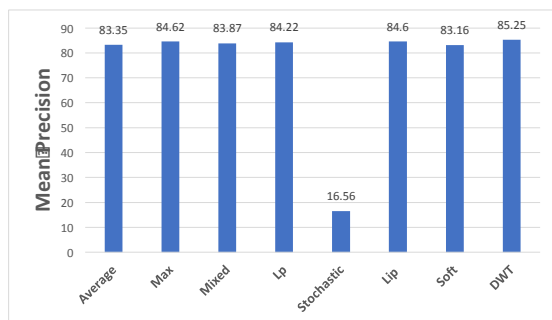
۲. Accuracy.

استاندارد است. شکل ۱۰ نشان می‌دهد که VGG۱۶ با BN بهترین نتایج را با استفاده ادغام ماکسیمم به دست 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 مکرر 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).

|                 |                                                                  |
|-----------------|------------------------------------------------------------------|
| FP <sup>۱</sup> | تعداد داده‌های مثبتی که به اشتباه به عنوان منفی شناسایی شده‌اند. |
| TN <sup>۲</sup> | تعداد داده‌های منفی که به درستی منفی شناسایی شده‌اند.            |
| FN <sup>۳</sup> | تعداد داده‌های منفی که به اشتباه مثبت تشخیص داده شده‌اند.        |

صحت تشخیص، نسبت کلاس‌های پیش‌بینی شده درست به کل داده‌ها است. از نظر ریاضی می‌توان آن را به صورت زیر بیان کرد:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (۱۰)$$

۴.۳.۲. یادآوری

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

$$Recall = \frac{TP}{TP+FN} \quad (۱۱)$$

۴.۳.۳. دقت

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:

$$Precision = \frac{TP}{TP+FP} \quad (۱۲)$$

۴.۳.۴. کاپا کوهن

ضریب کاپا کوهن مقداری در  $[-1, +1]$  است و هر چه به  $+1$  نزدیک‌تر باشد، توافقی<sup>۷</sup> متناسب‌تر و مستقیم‌تر است. مقدار نزدیک  $-1$  نشان دهنده توافق معکوس است، در حالی که یک نزدیک به صفر، نشان‌دهنده عدم توافق است. معادله ۱۳ ضریب  $[52]$  را محاسبه می‌کند.

$$Kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \quad (۱۳)$$

به‌طوری که  $Pr(a)$  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

۵. Precision.

۶. Cohen Kappa.

۷. Agreement.

۱. False Positive: FP.

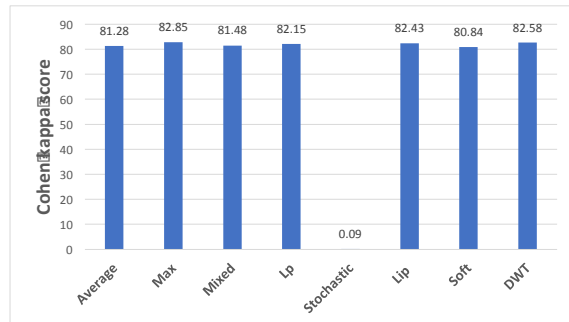
۲. True Negative: TN.

۳. False Negative: FN.

۴. Recall.



مجموعه داده را مدیریت می‌کند. می‌توان گفت که این یک اندازه-گیری، 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.

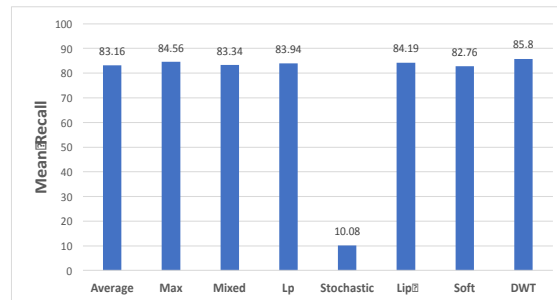


شکل ۱۳: مقایسه معیار کاپا کوهن مدل‌های

این مقاله از شبکه عصبی VGG<sup>۱۶</sup> برای اندازه‌گیری عملکرد روش‌ها استفاده کرده است. به دلیل ساختار عمیق و تعداد بالای عناصر قابل آموزش، آموزش آن تنها با نرمال‌سازی دسته‌ای امکان‌پذیر است. همان‌طور که در جدول ۲ نشان داده شده است، لایه‌های ادغام مختلف، زمان‌های شناسایی مختلفی را برای VGG<sup>۱۶</sup> 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<sup>۱</sup> را در مرحله آموزش 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      | ۰/۵۹۲۴     | ۸۳/۱۶         | ۰/۶۳۰۹    | ۸۳/۲۵   | ۸۳/۳۵          | ۸۳/۱۶       | ۸۱/۲۸             | ۰/۲۹۶ |
| Max          | ۰/۵۱۶۸     | ۸۴/۵۷         | ۰/۵۹۳۰    | ۸۴/۵۸   | ۸۴/۶۲          | ۸۴/۵۶       | ۸۲/۸۵             | ۰/۲۹۱ |
| Mixed        | ۰/۵۰۴۳     | ۸۳/۳۴         | ۰/۶۶۵۴    | ۸۳/۶۰   | ۸۳/۸۷          | ۸۳/۳۴       | ۸۱/۴۸             | ۰/۳۱۸ |

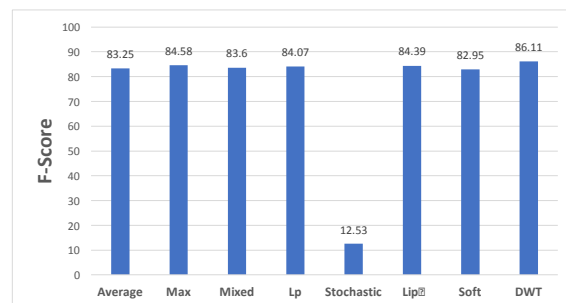


شکل ۱۱: مقایسه میانگین فراخوانی مدل‌های ادغام.

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

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

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

<sup>۱</sup> Loss

|            |        |       |        |       |       |       |       |       |
|------------|--------|-------|--------|-------|-------|-------|-------|-------|
| Lp         | ۰/۵۳۰۷ | ۸۳/۹۴ | ۰/۶۰۴۷ | ۸۴/۰۷ | ۸۴/۲۲ | ۸۳/۹۴ | ۸۲/۱۵ | ۰/۳۱۴ |
| Stochastic | ۱/۸۲۵۶ | ۱۰/۰۹ | ۲/۳۰۸۸ | ۱۲/۵۳ | ۱۶/۵۶ | ۱۰/۰۸ | ۰/۰۹  | ۰/۷۹۵ |
| Lip        | ۰/۵۳۲۹ | ۸۴/۱۹ | ۰/۶۱۱۶ | ۸۴/۳۹ | ۸۴/۶۰ | ۸۴/۱۹ | ۸۲/۴۳ | ۰/۴۳۵ |
| Soft       | ۰/۵۶۸۶ | ۸۲/۷۶ | ۰/۶۹۸۷ | ۸۲/۹۵ | ۸۳/۱۶ | ۸۲/۷۶ | ۸۰/۸۴ | ۰/۳۳۱ |
| DWT        | ۰/۲۶۸۵ | ۸۵/۶۷ | ۰/۳۵۶۸ | ۸۶/۱۱ | ۸۵/۲۵ | ۸۵/۸۰ | ۸۲/۵۸ | ۰/۵۲۴ |

## ۶. نتیجه‌گیری

این مقاله، به بررسی روش‌های مختلف ادغام محبوب برای مسائل و کاربردهای طبقه‌بندی پرداخته است. سپس، توانایی روش‌های ادغام را در ساختار VGG۱۶ در مجموعه داده CIFAR-۱۰ بررسی کرده است. این مطالعه شبکه VGG۱۶ را برای پارامترهای معیارهای طبقه‌بندی 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

[۸] 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. ۳۱, no. ۳, pp. ۲۷۵-۲۸۵, ۲۰۲۳.

[۹] N. Alturayef, H. Luqman, and M. Ahmed, "A systematic review of machine learning techniques for stance detection and its applications," *Neural Computing and Applications*, vol. ۳۵, no. ۷, pp. ۵۱۱۳-۵۱۴۴, ۲۰۲۳.

[۱۰] 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. ۱۱۶, no. ۴۴, pp. ۲۲۰۷۱-۲۲۰۸۰, ۲۰۱۹.

[۱۱] C. F. F. Sousa, "Categories' churn: a machine learning approach in retail," ۲۰۲۳.

[۱۲] 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. ۳-۱۷, ۲۰۲۱.

[۱۳] K. A. Taher, B. M. Y. Jisan, and M. M. Rahman, "Network intrusion detection using supervised machine learning technique with feature selection," in ۲۰۱۹ International conference on robotics, electrical and signal processing techniques (ICREST), IEEE, pp. ۶۴۳-۶۴۶, ۲۰۱۹.

[۱۴] L. Rundo et al, "Automated prostate gland segmentation based on an unsupervised fuzzy C-means clustering technique using multispectral T۱w and T۲w MR imaging," *Information*, vol. ۸, no. ۲, p. ۴۹, ۲۰۱۷.

[۱۵] 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. ۲۶, no. ۱, pp. ۷۹-۸۷, ۲۰۲۱.

[۱۶] 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. ۵۴, pp. ۲۸۰-۲۹۶, ۲۰۱۹.

[۱۷] D. Kangin and N. Pugeault, "Combination of Supervised and Reinforcement Learning For Vision-Based Autonomous Control," ۲۰۱۸.

[۱۸] 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 ۲۰۱۹ IEEE International Conference on Real-time Computing and Robotics (RCAR), IEEE, pp. ۱۱۲-۱۱۷, ۲۰۱۹.

[۱] Z. Lu et al., "Multiobjective Evolutionary Design of Deep Convolutional Neural Networks for Image Classification," *IEEE Transactions on Evolutionary Computation*, vol. ۲۵, no. ۲, pp. ۲۷۷-۲۹۱, ۲۰۲۰.

[۲] 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. ۹, no. ۱, pp. ۱-۹, ۲۰۲۱.

[۳] A. Pal et al., "Deep Metric Learning for Cervical Image Classification," *IEEE Access*, vol. ۹, pp. ۵۳۲۶۶-۵۳۲۷۵, ۲۰۲۱.

[۴] 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. ۲۳۸, p. ۱۲۲۱۱۱, ۲۰۲۴.

[۵] 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. ۱۴۱, pp. ۶۱-۶۷, ۲۰۲۱.

[۶] 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. ۵۵۸-۵۶۷, ۲۰۱۹.

[۷] H. A. Alsattar et al, "Developing deep transfer and machine learning models of chest X-ray for diagnosing COVID-۱۹ cases using probabilistic single-valued neutrosophic hesitant fuzzy," *Expert Systems with Applications*, vol. ۲۳۶, p. ۱۲۱۳۰۰, ۲۰۲۴.

- [۳۲] M. Imani and H. Ghassemian, "An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges," *Information fusion*, vol. ۵۹, pp. ۵۹-۸۳, ۲۰۲۰.
- [۳۳] 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. ۹۹-۱۱۰, ۲۰۲۰.
- [۳۴] 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 Classification," in *International Conference on Intelligent Systems Design and Applications*, pp. ۱۷۱-۱۸۰, ۲۰۲۰.
- [۳۵] V. Christlein, L. Spranger, M. Seuret, A. Nicolaou, P. Král, and A. Maier, "Deep generalized max pooling," in *۲۰۱۹ International Conference on Document Analysis and Recognition (ICDAR)*, pp. ۱۰۹۰-۱۰۹۶, ۲۰۱۹.
- [۳۶] T. Sharma, N. K. Verma, and S. Masood, "Mixed fuzzy pooling in convolutional neural networks for image classification," *Multimedia Tools and Applications*, vol. ۸۲, no. ۶, pp. ۸۴۰۵-۸۴۲۱, ۲۰۲۳.
- [۳۷] 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. ۳۶۴-۳۷۵, ۲۰۱۴.
- [۳۸] 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. ۸۲, no. ۷, pp. ۹۶۲۷-۹۶۴۸, ۲۰۲۳.
- [۳۹] P. Msonda, S. A. Uymaz, and S. S. Karaağaç, "Spatial Pyramid Pooling in Deep Convolutional Networks for Automatic Tuberculosis Diagnosis," *Traitement du Signal*, vol. ۳۷, no. ۶, ۲۰۲۰.
- [۴۰] 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. ۱۴, no. ۷, pp. ۸۱۳۵-۸۱۵۲, ۲۰۲۳.
- [۴۱] 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. ۳۷, no. ۹, pp. ۱۹۰۴-۱۹۱۶, ۲۰۱۵.
- [۴۲] J. Zhou, Y. Tian, C. Yuan, K. Yin, G. Yang, and M. Wen, "Improved uav opium poppy detection using an updated yolov۳ model," *Sensors*, vol. ۱۹, no. ۲۲, p. ۴۸۵۱, ۲۰۱۹.
- [۴۳] I. Hwang and K. Lee, "Removing Speaker Information from Speech Representation using Variable-Length Soft Pooling," *arXiv preprint arXiv:۲۴۰۴.۰۰۸۵۶*, ۲۰۲۴.
- [۴۴] A. Stergiou, R. Poppe, and G. Kalliatakis, "Refining activation downsampling with Softpool," *arXiv preprint arXiv:۲۱۰۱.۰۰۴۴۰*, ۲۰۲۱.
- [۱۹] 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. ۱۷۸, pp. ۴۱۴-۴۲۳, ۲۰۲۰.
- [۲۰] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," *Progress in Artificial Intelligence*, vol. ۹, no. ۲, pp. ۸۵-۱۱۲, ۲۰۲۰.
- [۲۱] 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 ۲nd International Conference on Artificial Intelligence and Speech Technology (AIST۲۰۲۰)*, ۱۹-۲۰ November, ۲۰۲۰, Delhi, India, CRC Press, p. ۱۲۱, ۲۰۲۱.
- [۲۲] P. P. Liang, A. Zadeh, and L.-P. Morency, "Foundations & trends in multimodal machine learning: Principles, challenges, and open questions," *ACM Computing Surveys*, vol. ۵۶, no. ۱۰, pp. ۱-۴۲, ۲۰۲۴.
- [۲۳] 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. ۶۵, no. ۴, pp. ۱۳۴۹-۱۳۶۲, ۲۰۱۷.
- [۲۴] 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. ۱۱۵۴۲۰, ۲۰۲۱.
- [۲۵] 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. ۲۲۴-۲۳۶, ۲۰۱۹.
- [۲۶] 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. ۴۰, no. ۴, pp. ۸۳۴-۸۴۸, ۲۰۱۷.
- [۲۷] J. Jordan, "Common architectures in convolutional neural networks," *Режим доступу: <https://www.jeremyjordan.me/convnet-architectures>*, ۲۰۱۸.
- [۲۸] Z. He, "Deep Learning in Image Classification: A Survey Report," in *۲۰۲۰ ۲nd International Conference on Information Technology and Computer Application (ITCA)*, , pp. ۱۷۴-۱۷۷, ۲۰۲۰.
- [۲۹] 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. ۶۲, no. ۱, pp. ۳۹۹-۴۱۱, ۲۰۲۰.
- [۳۰] H. Gholamalinezhad and H. Khosravi, "Pooling methods in deep neural networks, a review," *arXiv preprint arXiv:۲۰۰۹.۰۷۴۸۵*, ۲۰۲۰.
- [۳۱] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. ۱۸۷, pp. ۲۷-۴۸, ۲۰۱۶.

- [۴۵] 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. ۱۸۳, p. ۱۱۵۴۲۰, ۲۰۲۱.
- [۴۶] 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. ۱۷, no. ۱, p. ۰۱۶۰۷۳, ۲۰۲۰.
- [۴۷] W. Cukierski, "CIFAR-۱۰-Object Recognition in Images," Kaggle ۲۰۱۳ Available online: <https://www.kaggle.com/c/cifar-10> (accessed on ۳۰ December ۲۰۲۲), ۲۰۱۳.
- [۴۸] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," ۲۰۰۹.
- [۴۹] 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*, ۲۰۱۸.
- [۵۰] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:۱۴۰۹.۱۵۵۶*, ۲۰۱۴.
- [۵۱] T. Sulistyowati, P. PURWANTO, F. Alzami, and R. A. Pramunendar, "Vgg۱۶ deep learning architecture using imbalance data methods for the detection of apple leaf diseases," *Moneter: Jurnal Keuangan dan Perbankan*, vol. ۱۱, no. ۱, pp. ۴۱-۵۳, ۲۰۲۳.
- [۵۲] 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. ۱-۸, ۲۰۱۰.

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

Azam Noroozi<sup>۱\*</sup>

<sup>۱</sup>- 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-۱۰ 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.