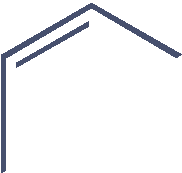
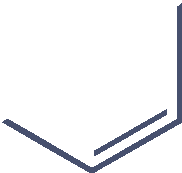
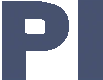
[از راه دور حسگری](https://www.mdpi.com/journal/remotesensing)



*مقاله*

**چشم انداز ترانسفورماتورها برای از راه دور سنجش تصویر طبقه‌بندی**

# یعقوب بازی ۱، \* ، لیلا باشمال ۱ ، محمد م. آل راهال ۲ ، رهام آل دییل ۱ و نایف آل عجلان ۱

۱ گروه مهندسی کامپیوتر، دانشکده علوم کامپیوتر و اطلاعات، دانشگاه ملک سعود، ریاض 11543، عربستان سعودی; [439204359@student.ksu.edu.sa](mailto:439204359@student.ksu.edu.sa) (LB); [437204127@student.ksu.edu.sa](mailto:437204127@student.ksu.edu.sa) (RAD); [najlan@ksu.edu.sa](mailto:najlan@ksu.edu.sa) (NAA)

۲ کاربردی کامپیوتر علم دپارتمان، کالج از کاربردی کامپیوتر علم، پادشاه سعود دانشگاه،

ریاض ۱۱۵۴۳، سعودی عربستان سعودی؛ [mmalrahhal@ksu.edu.sa](mailto:mmalrahhal@ksu.edu.sa)

**\*** مکاتبات: [ybazi@ksu.edu.sa;](mailto:ybazi@ksu.edu.sa) تلفن: +966- 101469629

[بررسی رور](https://www.mdpi.com/2072-4292/13/3/516?type=check_update&version=3)

[**به‌روزرسانی‌ها**](https://www.mdpi.com/2072-4292/13/3/516?type=check_update&version=3)

**استناد:** بزی، ی.؛ باشمال، ل.؛ رحال، ام ام ای؛ دییل، آر ای؛ عجلان،

نه چشم انداز ترانسفورماتورها برای از راه دور سنجش تصویر طبقه‌بندی. *از راه دور سنس* **۲۰۲۱** ، *۱۳* ، ۵۱۶. [https://doi.org/](https://doi.org/10.3390/rs13030516) [۱۰.۳۳۹۰/rs13030516](https://doi.org/10.3390/rs13030516)

ویراستار علمی: چی وانگ دریافت شده: ۲۳ دسامبر ۲۰۲۰

پذیرفته شده: ۲۹ ژانویه ۲۰۲۱

منتشر شده: ۱ فوریه ۲۰۲۱

**یادداشت ناشر:** MDPI بی‌طرف می‌ماند در خصوص دعاوی قضایی در نقشه‌های منتشر شده و وابستگی‌های نهادی عملیات

[](https://creativecommons.org/)

**حق نشر:** © ۲۰۲۱ توسط ‎‏ ... نویسندگان. دارنده مجوز ام دی پی آی، بازل، سوئیس. این مقاله است یک باز دسترسی مقاله توزیع شده تحت ‎‏ ... اصطلاحات و شرایط از ‎‏ ... خلاق عوام انتساب (سی سی توسط) مجوز [(https://](https://creativecommons.org/licenses/by/4.0/) [creativecommons.org/licenses/by/](https://creativecommons.org/licenses/by/4.0/) ۴.۰/).

**چکیده:** در این مقاله، ما یک روش طبقه‌بندی صحنه سنجش از دور مبتنی بر بینایی پیشنهاد می‌کنیم. ترانسفورماتورها اینها انواع از شبکه‌ها، که هستند الان به رسمیت شناخته شده به عنوان مدرن مدل‌ها در طبیعی پردازش زبان، مانند شبکه‌های عصبی کانولوشن استاندارد، به لایه‌های کانولوشن متکی نیستند. (سی‌ان‌ان‌ها). در عوض، آنها استفاده چند سر توجه مکانیسم‌ها به عنوان ‎‏ ... اصلی ساختمان بلوک به مشتق شده رابطه زمینه‌ای دوربرد بین پیکسل‌ها در تصاویر. در مرحله اول، تصاویر مورد تجزیه و تحلیل قرار می‌گیرند به تکه‌هایی تقسیم شده، سپس با مسطح‌سازی و جاسازی به توالی تبدیل می‌شوند. برای حفظ اطلاعات درباره ‎‏ ... موقعیت، تعبیه کردن موقعیت است اضافه شد به اینها وصله‌ها سپس، ‎‏ ... در نتیجه توالی است برای تولید نمایش نهایی به چندین لایه توجه چند وجهی داده می‌شود. در طبقه‌بندی در مرحله، اولین توالی توکن به یک لایه طبقه‌بندی softmax داده می‌شود. برای افزایش رتبه‌بندی عملکرد، ما کاوش کردن چندین داده‌ها تقویت استراتژی‌ها به تولید کردن اضافی داده‌ها برای آموزش. علاوه بر این، ما به صورت تجربی نشان می‌دهیم که می‌توانیم با هرس کردن نیمی از لایه‌ها، شبکه را فشرده کنیم. در عین حال، دقت طبقه‌بندی رقابتی را حفظ می‌کند. نتایج آزمایش‌های انجام‌شده روی نمونه‌های مختلف سنجش از دور تصویر مجموعه داده‌ها نشان دادن ‎‏ ... نویدبخش قابلیت از ‎‏ ... مدل مقایسه شده به روش‌های پیشرفته. به طور خاص، Vision Transformer دقت طبقه‌بندی متوسطی را به دست می‌آورد. به ترتیب ۹۸.۴۹٪، ۹۵.۸۶٪، ۹۵.۵۶٪ و ۹۳.۸۳٪ روی مجموعه داده‌های Merced، AID، Optimal31 و NWPU. در حالی که نسخه فشرده شده به دست آمده با حذف نیمی از لایه‌های توجه چند وجهی، نتیجه می‌دهد به ترتیب ۹۷.۹۰٪، ۹۴.۲۷٪، ۹۵.۳۰٪ و ۹۳.۰۵٪.

**کلمات کلیدی:** سنجش از دور؛ طبقه‌بندی در سطح تصویر؛ مبدل‌های بینایی؛ توجه چندوجهی؛ داده‌افزایی

# مقدمه

سنجش از دور (RS) علم جمع‌آوری اطلاعات در مورد اشیاء بدون هیچ گونه تماس فیزیکی مستقیم، معمولاً از طریق ماهواره، هواپیما یا وسیله نقلیه هوایی بدون سرنشین است. (پهپاد) [ [1](#_bookmark17) ]. مثال‌ها از کاربردها از از راه دور حسگری شامل شدن زمین شناسی نقشه برداری، آزمایش محیط زیست، اکتشاف نفت، مدیریت ترافیک، پیش بینی زلزله و ساخت و ساز در زمینه حفاظت از آب [ [2](#_bookmark18) ، [3](#_bookmark19) ].

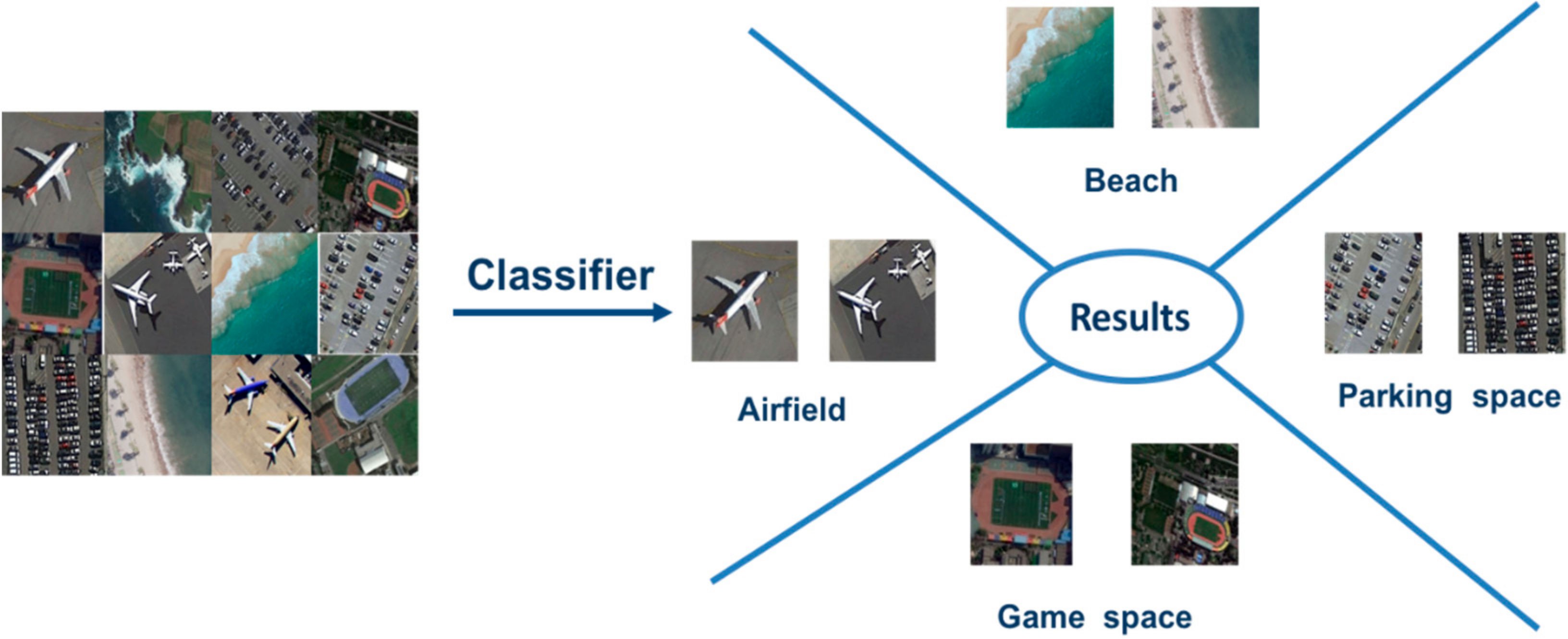
تکامل حسگرهای ماهواره‌ای، از نظر وضوح مکانی و زمانی بهبود یافته‌اند که فرصت‌هایی را برای تفکیک جزئیات دقیق فراهم می‌کند. ‎‏ ... زمین سطح. ماهواره‌ها چنین به عنوان مودیس (1 کیلومتر ۱ کیلومتر) ارائه حرارتی داده‌ها با وضوح زمانی بالا، از وضوح مکانی پایین رنج می‌برند. از سوی دیگر، لندست تغییرات در مقیاس کوچک ۱۰۰ تا ۲۰۰ متر را با وضوح زمانی بسیار پایین ارائه می‌دهد. نسل جدید ماهواره‌ها می‌توانند تصاویر طیفی و مکانی بسیار بالایی ارائه دهند؛ برای مثال، IKONOS-2 تولید می‌کند تصاویر با ۴ باند چند طیفی وضوح تصویر و فضایی وضوح تصویر از 2.5 تا 4 متر. وسایل نقلیه هوایی بدون سرنشین (UAV) یک راه حل بهبود یافته از پلتفرم‌های سنجش از دور را ارائه می‌دهند که در سال‌های گذشته شاهد رشد بالایی بوده‌اند و به طور گسترده برای تشخیص آتش‌سوزی، نقشه‌برداری نظارتی و نظارت بر رانش زمین و موارد دیگر استفاده می‌شوند. استفاده می‌کند [ [4](#_bookmark20) ]. پهپادها داشته باشند چندین مزایا بیش از ماهواره و هوایی تصاویر. اول، آنها هستند

×

*از راه دور سنس* **۲۰۲۱** ، *۱۳* ، ۵۱۶. <https://doi.org/10.3390/rs13030516> <https://www.mdpi.com/journal/remotesensing>

استقرار آسان‌تر برای برآوردن الزامات نظارت، ارزیابی و نقشه‌برداری سریع. آنها می‌تواند کار در پایین‌تر ارتفاعات مقایسه شده به ‎‏ ... خلبانی شده هواپیما، که فراهم می‌کند وضوح مکانی در ‎‏ ... سانتی‌متر سطح. آنها می‌تواند پرواز هر زمان ‎‏ ... آب و هوا مجوزها، پیشرو به بهبود در وضوح زمانی. با افزایش وضوح مکانی، تصاویر احتمالاً حاوی توصیفگرهای نویزی و پرت هستند. یک مطالعه اخیر [ [5](#_bookmark21) ] یک شبکه عصبی کانولوشن (CNN) را برای طبقه‌بندی تصاویر گرفته شده توسط دوربین نصب شده بر روی یک پهپاد پیشنهاد کرد. النجار و همکارانش [ [6](#_bookmark22) ] یک مدل CNN را برای طبقه‌بندی یک مدل سطح دیجیتال در کنار تصاویر پهپاد پیشنهاد کردند. لیو و همکارانش [ [7](#_bookmark23) ] شبکه‌های عصبی کانولوشن را با تجزیه و تحلیل تصویر مبتنی بر شیء (OBIA) برای پوشش زمین ترکیب کردند. طبقه بندی با استفاده از چندنمایی داده‌ها نویسنده در [ [8](#_bookmark24) ] پیشنهادی الف دو شاخه شبکه عصبی برای اختصاص برچسب‌های کلاس چندگانه به تصاویر پهپاد.

موضوع طبقه‌بندی صحنه اخیراً به یک حوزه تحقیقاتی فعال تبدیل شده است تا با مشکل چالش‌برانگیز تفسیر مؤثر تصاویر سنجش از دور روبرو شود. وظیفه گرفتن است یک تصویر و به درستی برچسب زدن آن به الف از پیش تعریف شده کلاس به عنوان نشان داده شده در شکل [۱. (۱](#_bookmark0) ) طبقه‌بندی صحنه یک کار مهم برای بسیاری از کاربردها، مانند مدیریت زمین [ [9](#_bookmark25) ]، برنامه‌ریزی شهری [ [10](#_bookmark26) ] و مدل‌سازی آتش‌سوزی‌های جنگلی [ [11](#_bookmark27) ] است.



**شکل ۱.** از راه دور حسگری صحنه طبقه بندی.

کارهای اولیه در زمینه طبقه‌بندی صحنه بر اساس ویژگی‌های دست‌ساز، استخراج دستی توسط انسان، از جمله الگوهای دودویی محلی (LBP) [ [12](#_bookmark28) ]، هیستوگرام گرادیان‌های جهت‌دار (HOG) [ [13](#_bookmark29) ] و تبدیل ویژگی مقیاس‌ناپذیر (SIFT) [ [14](#_bookmark30) ] بود. روش‌های مرسوم طبقه‌بندی صحنه به رمزگذاری ویژگی‌های دست‌ساز با مدل‌های مختلف مانند کیسه کلمات (BoWs) [ [15](#_bookmark31) ]، بردارهای فیشر (FV) [ [16](#_bookmark32) ] یا بردار توصیفگرهای تجمیع‌شده محلی (VLAD) [ [17](#_bookmark33) ] بستگی دارند.

روشن ‎‏ ... دیگر دست، عمیق یادگیری روش‌ها چنین به عنوان عمیق باور شبکه‌ها (DBN ها) [ [18](#_bookmark34) ] و رمزگذارهای خودکار انباشته [ [19](#_bookmark35) ] در چندین کاربرد، از جمله موارد زیر، دستاوردهای عظیمی کسب کردند. سنجش از دور تصویر طبقه‌بندی. در خاص، سی‌ان‌ان‌ها داشته باشند پیشی گرفته سنتی روش‌ها در بسیاری کاربردها [ [20-22](#_bookmark36) ] . [اینها](#_bookmark37) روش‌ها داشته باشند الف اصلی کلید مزیت ارائه یک راه‌حل جامع که به حداقل مهندسی ویژگی نیاز دارد. رویکردهای دیگر مبتنی بر روی مکرر عصبی شبکه‌ها (RNN ها) [ [23](#_bookmark38) ]، مولد خصمانه شبکه‌های عصبی مصنوعی (GAN) [ [24](#_bookmark39) ، [25](#_bookmark40) ]، شبکه‌های کانولوشن گراف (GCN) [ [26](#_bookmark41) ] و حافظه بلندمدت-کوتاه‌مدت (LSTM) [ [27](#_bookmark42) ] نیز معرفی شده‌اند. در یک مقاله اخیر، نویسندگان طبقه‌بندی صحنه‌های سنجش از دور را به عنوان یک مسئله یادگیری چند نمونه‌ای (MIL) در نظر گرفتند [ [28](#_bookmark43) ]. آنها یک شبکه چند نمونه‌ای با اتصال متراکم را برای برجسته کردن معانی محلی مرتبط با برچسب صحنه پیشنهاد کردند. این روش، توانایی محلی را افزایش می‌دهد. معنایی نمایندگی توسط به طور موثر دور انداختن بی فایده اطلاعات یو و ال. [ [29](#_bookmark44) ] شبکه مولد تخاصمی توجه (Attention GAN) را پیشنهاد کرد که شبکه‌های مولد تخاصمی را با مکانیسم توجه ادغام می‌کند تا توجه را افزایش دهد. ‎‏ ... نمایندگی قدرت از ‎‏ ... تبعیض کننده برای هوایی صحنه طبقه بندی. نویسندگان در [ [30](#_bookmark45) ] یک روش تنظیم دقیق ساده با استفاده از یک خطای طبقه‌بندی کمکی معرفی کردند. آنها نشان داد چگونه به مبارزه ‎‏ ... ناپدید شدن گرادیان مشکل با استفاده از یک کمکی تابع زیان. سان و همکارانش [ [31](#_bookmark46) ] یک شبکه دوطرفه دروازه‌دار برای ادغام ویژگی‌ها پیشنهاد کردند. لیو

و همکاران [ [32](#_bookmark47) ] نقشه‌های ویژگی را از لایه‌های میانی و کاملاً متصل و ورودی ترکیب کردند. آنها به ‎‏ ... طبقه بندی کننده برای طبقه بندی. یو و ال. [ [33](#_bookmark48) ] ترکیب شده دو از پیش آموزش دیده CNN ها با ‎‏ ... دو جریانی همجوشی تکنیک به طبقه بندی کردن با وضوح بالا هوایی صحنه‌ها چنگ و همکارانش [ [34](#_bookmark49) ] یک تنظیم یادگیری متریک بر روی ویژگی‌های متمایزکننده CNNها برای بهینه‌سازی پیشنهاد کردند. الف جدید تبعیض آمیز عینی تابع به ساختن ‎‏ ... مدل بیشتر Xue و همکارانش [ [35](#_bookmark50) ] روشی را با استفاده از سه شبکه عمیق برای استخراج ویژگی‌های عمیق از تصویر به صورت جداگانه پیشنهاد کردند. سپس این ویژگی‌ها با هم ترکیب شدند تا یک بردار ویژگی واحد برای طبقه‌بندی ایجاد شود.

علاوه بر CNNها، نوع جدیدی از مدل‌های یادگیری عمیق به نام Transformers پیشنهاد شده و در بینایی کامپیوتر محبوبیت خاصی پیدا کرده‌اند. Transformers به یک روش ساده اما ... متکی هستند. قدرتمند رویه نامیده می شود توجه، که تمرکزها روی قطعی قطعات از ‎‏ ... ورودی برای دستیابی به نتایج کارآمدتر. در حال حاضر، آنها به عنوان مدل‌های پیشرفته در داده‌های ترتیبی، به ویژه روش‌های پردازش زبان طبیعی (NLP) مانند ترجمه ماشینی [ [36](#_bookmark51) ]، مدل‌سازی زبان [ [37](#_bookmark52) ] و تشخیص گفتار [ [38](#_bookmark53) ] در نظر گرفته می‌شوند. معماری ترانسفورماتور توسعه‌یافته توسط واسوانی و همکارانش [ [39](#_bookmark54) ] بر اساس مدل رمزگذار-رمزگشا است، که دگرگون می‌کند الف داده شده توالی از عناصر به دیگری دنباله. انگیزه اصلی برای ترانسفورماتورها بود به فعال کردن موازی پردازش از ‎‏ ... کلمات در الف جمله، که در LSTMها یا RNNها امکان‌پذیر نبود زیرا آنها کلمات یک جمله را می‌گیرند توسط یکی.

با الهام از موفقیت Transformers در پردازش زبان طبیعی (NLP)، تحقیقات جدید سعی در اعمال مستقیم Transformers بر روی تصاویر دارند. این یک کار چالش برانگیز است، به دلیل نیاز به کاربرد خود-توجهی که در آن هر پیکسل به تمام پیکسل‌های دیگر توجه می‌کند. برای تصاویر، این کار بسیار پرهزینه است زیرا تصویر شامل تعداد زیادی پیکسل است. محققان چندین رویکرد را برای اعمال Transformers به تصاویر امتحان کردند. برخی از آثار، معماری‌های CNN را با خود-توجهی ترکیب کردند. برای مثال، بلو و ال. [ [40](#_bookmark55) ] تقویت‌شده سی‌ان‌ان‌ها توسط جایگزین کردن برخی لایه‌های کانولوشن با توجه به خود لایه‌ها، که رهبری کرد به بهبود در تصویر طبقه‌بندی. با این حال ، این روش با هزینه محاسباتی بالایی مواجه بود زیرا اندازه بزرگ تصویر باعث ... می‌شود. عظیم رشد در ‎‏ ... زمان پیچیدگی از توجه به خود. وانگ و ال. [ [41](#_bookmark56) ] پیشنهادی یک روش به تولید کردن قدرتمند ویژگی‌ها توسط به طور گزینشی تمرکز کردن روی بحرانی قطعات یا مکان‌های تصویر، و سپس پردازش متوالی آنها. وو و همکارانش [ [42](#_bookmark57) ] از ترانسفورماتور روی ... استفاده کردند. ‎‏ ... سی‌ان‌ان؛ اول آنها استخراج شده ویژگی نقشه‌ها با استفاده از الف سی‌ان‌ان، سپس تغذیه شده آنها به انباشته شده ترانسفورماتورهای بصری به فرآیند بصری توکن ها و محاسبه ‎‏ ... خروجی. راماچاندران و ال. [ [43](#_bookmark58) ] اول شروع شده به استفاده توجه به خود به عنوان الف مستقل ساختمان بلوک برای بینایی وظایف در عوض از یک ساده تقویت روی بالا از کانولوشن لایه‌ها. آنها مجموعه بالا الف به طور کامل توجه مدل توسط جایگزین کردن همه کانولوشن لایه‌ها با توجه به خود لایه‌ها چن و ال. [ [44](#_bookmark59) ] پیشنهادی روشی که Transformers را روی تصاویر خام با وضوح کاهش‌یافته اعمال می‌کند و آن‌ها را به توالی‌های طولانی پیکسل‌های متن‌مانند تغییر شکل می‌دهد.

در الف خیلی اخیر مشارکت، و متفاوت از قبلی کار می‌کند، دوسوویتسکی و ال. [ [45](#_bookmark60) ] اعمال شده الف استاندارد ترانسفورماتور مستقیماً به تصاویر توسط تقسیم کردن ‎‏ ... تصویر به وصله‌ها نه تمرکز کردن روی پیکسل‌ها، سپس ورودی به ‎‏ ... ترانسفورماتور ‎‏ ... توالی از جاسازی‌ها برای اون وصله‌ها. تکه‌های تصویر در برنامه‌های NLP به عنوان توکن در نظر گرفته شدند. اینها مدل‌ها منجر به نتایج بسیار رقابتی در مجموعه داده ImageNet شدند. در این کار، ما خواهیم از این مدل‌های از پیش آموزش‌دیده برای انتقال دانش به مورد تصاویر سنجش از دور استفاده کنید. در واقع، به ‎‏ ... بهترین از ما دانش در سنجش از دور در وظایف طبقه‌بندی صحنه، معماری‌های کانولوشن همچنان غالب هستند و ترانسفورماتورها هنوز به عنوان انتخاب مدل در طبقه‌بندی به طور گسترده مورد استفاده قرار نگرفته‌اند. به عنوان مثال، هی و همکارانش [ [46](#_bookmark61) ] مدلی مشتق شده از نمایش‌های رمزگذار دو جهته به نام BERT [ [47](#_bookmark62) ] را پیشنهاد کردند که در زمینه پردازش زبان طبیعی در زمینه تصاویر فراطیفی استفاده می‌شد. این روش است مبتنی بر روی چندین چند سر توجه به خود لایه‌ها هر کدام سر کدگذاری می‌کند ‎‏ ... آگاه از بافت معنایی نمایندگی به بدست آوردن تبعیض آمیز ویژگی‌ها که هستند مورد نیاز برای طبقه‌بندی دقیق در سطح پیکسل

در این مقاله، ما ارزیابی گسترده‌ای از مدل ارائه شده در [ [45](#_bookmark60) ] ارائه می‌دهیم. برای طبقه‌بندی تصاویر سنجش از دور برای این منظور، تصاویر مورد تجزیه و تحلیل تقسیم شده به وصله‌ها، سپس تبدیل شده به توالی توسط مسطح کردن و جاسازی

جاسازی موقعیت به این وصله‌ها اضافه می‌شود تا اطلاعات موقعیت را حفظ کند. به دست آمده توالی است سپس تغذیه شده به چندین چند سر توجه لایه‌ها برای مولد بازنمایی نهایی. در طول طبقه‌بندی، اولین توالی توکن به عنوان ورودی به یک لایه طبقه‌بندی softmax داده می‌شود. برای افزایش عملکرد طبقه‌بندی، چندین استراتژی افزایش داده مانند CutMix و Cutout را برای تولید داده‌های اضافی برای آموزش بررسی می‌کنیم. در ‎‏ ... آزمایش‌ها، ما نشان دادن که ما می‌تواند فشرده سازی ‎‏ ... شبکه توسط هرس کردن نیمی از لایه‌های آن را در حالی که دقت طبقه‌بندی رقابتی را حفظ می‌کند، پوشش می‌دهد.

ادامه مقاله به شرح زیر سازماندهی شده است: بخش [۲](#_bookmark1) روش‌های اصلی مبتنی بر ترانسفورماتورها را شرح می‌دهد. در بخش [۳](#_bookmark5) ، نتایج تجربی روی سه مجموعه داده شناخته شده را ارائه می‌دهیم. بخش [۴](#_bookmark13) بحثی در مورد نتایج ارائه می‌دهد و مقایسه‌هایی با روش‌های پیشرفته ارائه می‌دهد. سپس در نهایت در بخش [۵ نتیجه‌گیری می‌کنیم و مسیرهای آینده را نشان می‌دهیم](#_bookmark16) .

# مواد و روش‌ها

* 1. *چشم انداز ترانسفورماتور*

بگذارید *س* = { *ایکس آی* ، *y i* } *r* اهدا کردن الف مجموعه از *ر* از راه دور حسگری تصاویر، کجا *ایکس من* است یک تصویر

*i*=1

و *ی من* است آن متناظر کلاس برچسب *ی من* ۱، ۲، . . . ، *متر* ، و *متر* است ‎‏ ... شماره از تعریف شده

∈ { }

کلاس‌هایی برای آن مجموعه. هدف مدل Vision Transformer یادگیری نگاشت از توالی تکه‌های تصویر به برچسب معنایی مربوطه است.

Vision Transformer معماری‌ای است که کاملاً مبتنی بر Vanilla Trans- است. معماری سابق [ [39](#_bookmark54) ]، معماری‌ای که در سال‌های اخیر با نشان دادن عملکرد پیشرفته در ترجمه ماشینی و سایر وظایف پردازش زبان طبیعی [ [47](#_bookmark62) ] توجه زیادی را به خود جلب کرده است. دنبال می‌کند ‎‏ ... رمزگذار-رمزگشا معماری، با ‎‏ ... توانایی به فرآیند متوالی داده‌ها را به صورت موازی و بدون تکیه بر هیچ شبکه بازگشتی، منتقل می‌کند. موفقیت مدل‌های ترانسفورماتور تا حد زیادی بهره مند شد از ‎‏ ... توجه به خود مکانیسم، که است پیشنهادی به روابط بلندمدت بین عناصر توالی را ثبت کنید.

تبدیل‌کننده بینایی (Vision Transformer) به عنوان تلاشی برای گسترش استفاده از تبدیل‌کننده استاندارد در طبقه‌بندی تصویر پیشنهاد شده است. هدف اصلی، تعمیم آنها به روش‌هایی غیر از متن، بدون ادغام هیچ معماری داده-محور است. به طور خاص، Vision Transformer استفاده می‌کند ‎‏ ... رمزگذار ماژول از ‎‏ ... ترانسفورماتور به انجام دادن طبقه بندی با نقشه برداری الف توالی از تصویر وصله‌ها به ‎‏ ... معنایی برچسب. برخلاف ‎‏ ... متعارف در معماری‌های CNN که معمولاً از فیلترهایی با میدان دریافت محلی استفاده می‌کنند، مکانیزم توجه به کار رفته توسط Vision Transformer به آن اجازه می‌دهد تا در مناطق مختلف تصویر حضور داشته باشد و اطلاعات را در کل تصویر ادغام کند.

معماری کامل مدل از ابتدا تا انتها در شکل [2 نشان داده شده است](#_bookmark2) . به طور کلی، آن از یک لایه جاسازی، یک رمزگذار و یک طبقه‌بندی کننده نهایی سر تشکیل شده است. در مرحله اول، یک تصویر *X* از مجموعه آموزشی (برای سادگی، اندیس تصویر *i را حذف می‌کنیم* ) به تکه‌های غیرهمپوشان تقسیم می‌شود. هر تکه توسط Transformer به عنوان یک توکن مجزا در نظر گرفته می‌شود. بنابراین برای تصویری با اندازه *X، c ح w* (که در آن *h* ارتفاع، *w* عرض و *c* نشان دهنده ‎‏ ... شماره از کانال‌ها)، ما عصاره وصله‌ها هر کدام از بُعد *ج ص ص* از آن را.

× ×

× ×

این یک دنباله از تکه‌ها ( *x* 1 , ) را تشکیل می‌دهد. *ایکس* ۲ ، . . . ، *ایکس ان* ) به طول *n* ، با *n* = *hw* / *p^* 2 . معمولاً اندازه وصله *p* برابر با ۱۶ انتخاب می‌شود. ۱۶ یا ۳۲ ۳۲، که در آن اندازه پچ کوچکتر منجر به توالی طولانی‌تر می‌شود و برعکس.

× ×

* + 1. خطی جاسازی لایه

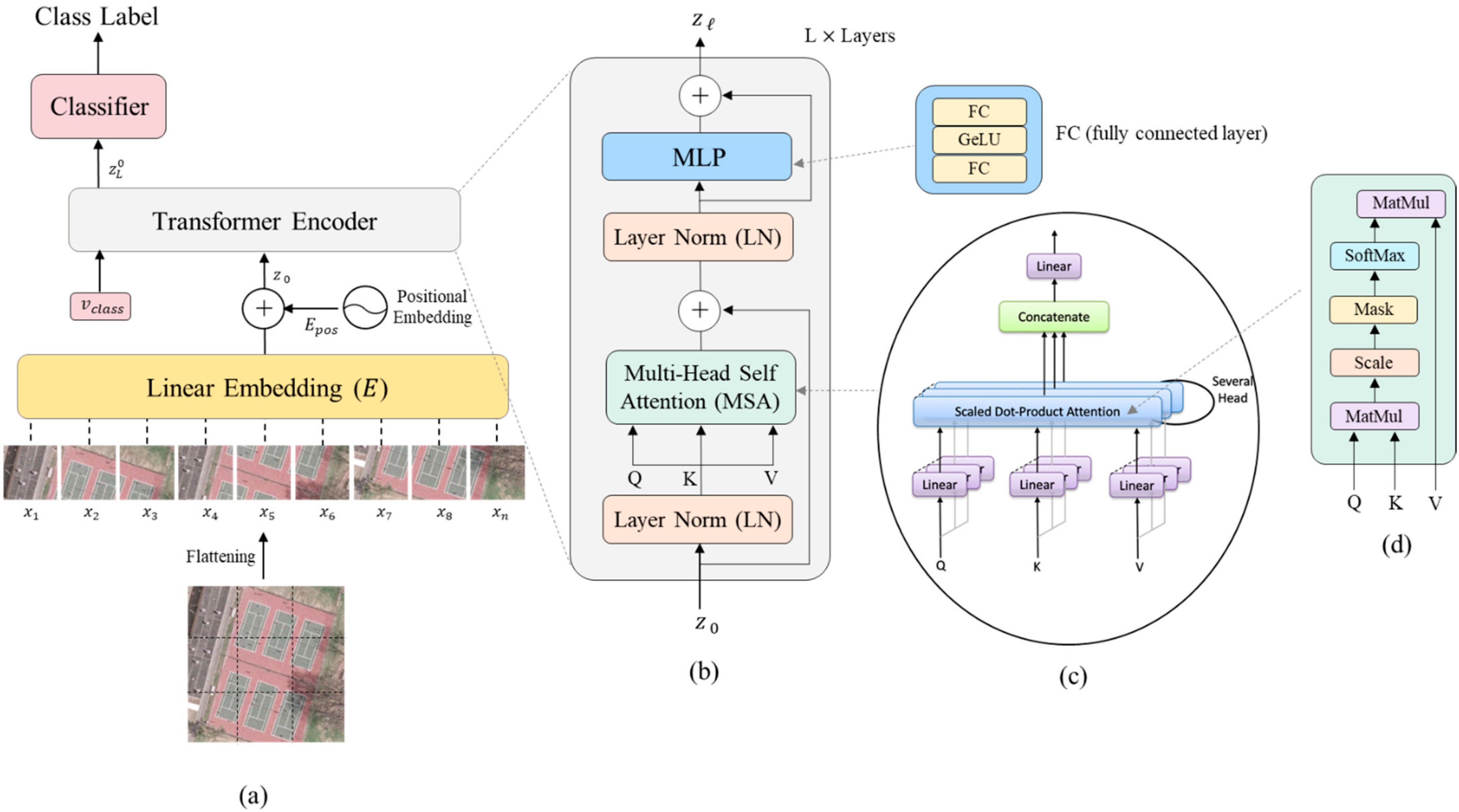
قبل از تغذیه ‎‏ ... توالی از وصله‌ها به ‎‏ ... رمزگذار، آن است به صورت خطی پیش‌بینی شده با استفاده از ماتریس جاسازی آموخته‌شده *E،* به برداری از بُعد مدل *d تبدیل می‌شود* . سپس نمایش‌های جاسازی‌شده به همراه یک *کلاس v توکن طبقه‌بندی قابل یادگیری به هم متصل می‌شوند.* که برای انجام وظیفه طبقه‌بندی مورد نیاز است. تکه‌های تصویر جاسازی‌شده مشاهده می‌شوند توسط ‎‏ ... ترانسفورماتور به عنوان الف مجموعه از وصله‌ها بدون هر مفهوم از آنها سفارش دهید. به نگه داشتن فضایی ترتیب از ‎‏ ... وصله‌ها به عنوان در ‎‏ ... اصلی تصویر، ‎‏ ... موقعیتی اطلاعات *پست الکترونیکی*

کدگذاری شده و به نمایش‌های پچ اضافه می‌شود. توالی جاسازی‌شده‌ی حاصل از وصله‌ها با توکن *z0* در (معادله (1)) داده شده است:

*ز* ۰ = [ *وی* ; *×* ۱ *E* ؛ *×* 2 *E* ; . . . ; *ایکس ان ای* ] + *پست الکترونیکی*​ *ای* ∈ ر ( *صفحه* ۲ *ج* ) × *د* ، *پست الکترونیکی* ∈ R ( *n* + 1 ) × *d* (1)

*class*

[45 ]](#_bookmark60) نشان داده شده است که کدگذاری‌های موقعیتی یک بعدی و دو بعدی نتایج تقریباً یکسانی تولید می‌کنند. بنابراین، از یک کدگذاری موقعیتی ساده‌ی یک‌بعدی برای حفظ اطلاعات موقعیتی تکه‌های مسطح‌شده استفاده می‌شود.



**شکل ۲.**  چشم انداز ترانسفورماتور معماری: ( **الف** ) ‎‏ ... اصلی معماری از ‎‏ ... مدل; ( **ب** ) ‎‏ ... ترانسفورماتور رمزگذار ماژول؛

( **ج** ) ‎‏ ... خود-چندمقیاسی توجه (MSA) سر، و ( **د** ) ‎‏ ... توجه به خود (SA) سر.

* + 1. چشم انداز ترانسفورماتور رمزگذار

توالی حاصل از وصله‌های جاسازی‌شده *z* 0 به رمزگذار ترانسفورماتور منتقل می‌شود. همانطور که در شکل [2b نشان داده شده است](#_bookmark2) ، رمزگذار از *L* لایه یکسان تشکیل شده است. هر کدام دو تا دارد زیرمولفه‌های اصلی: (1) یک بلوک خود-توجهی چندسره (MSA) (معادله (2)) و (2) یک بلوک متراکم پیش‌خور (MLP) کاملاً متصل (معادله (3))؛ بلوک دوم از دو لایه متراکم با یک فعال‌سازی GeLU در بین آنها تشکیل شده است. هر یک از دو زیرمولفه رمزگذار استخدام می‌کند باقیمانده رد شدن اتصالات و است مقدم توسط الف عادی سازی لایه (LN ) .

*z* ′ *l* = MSA ( LN ( *zl −* 1 ) ) + *z l* − 1 ، *ل* = ۱ . . . *ل* (2)

*ز ل* = MLP ال ان *z* ′ *l* + *z* ′ *l* ، *l* = 1 ... *L* (3) در ‎‏ ... آخرین لایه از ‎‏ ... رمزگذار، ما گرفتن ‎‏ ... اول عنصر در ‎‏ ... توالی *ز* و عبور

*L*

0

آن به یک خارجی سر طبقه بندی کننده برای پیش‌بینی ‎‏ ... کلاس برچسب.

*ی* = ال ان *ز* 0  (4)

*L*

بلوک MSA در انکودر، جزء مرکزی ترانسفورماتور است. این دارد نقش از تعیین کننده ‎‏ ... نسبی اهمیت از الف مجرد پچ تعبیه کردن با احترام به ‎‏ ... دیگر جاسازی‌ها در ‎‏ ... دنباله. این بلوک دارد چهار لایه‌ها: ‎‏ ... خطی لایه ،

لایه خود-توجه، لایه الحاق، که خروجی‌های چندین سر توجه را به هم متصل می‌کند، و یک لایه خطی نهایی، همانطور که در شکل 2c نشان داده شده [است](#_bookmark2) .

در سطح بالا، توجه را می‌توان با وزن توجه نشان داد که محاسبه می‌شود توسط پیدا کردن ‎‏ ... وزن دار جمع بیش از همه ارزش‌ها از ‎‏ ... توالی *ز* . توجه به خود سر (SA) با محاسبه مقیاس‌بندی پرس‌وجو-کلید-مقدار (پرس‌وجو-کلید-مقدار) وزن‌های توجه را یاد می‌گیرد. شکل [2d](#_bookmark2) جزئیات محاسباتی را که در بلوک SA انجام می‌شود نشان می‌دهد. برای هر عنصر در توالی ورودی، سه مقدار تولید می‌شود: *پرسش* (پرسش)، *ک* (کلید)، و *V* (مقدار) توسط ضرب کردن ‎‏ ... عنصر علیه سه یاد گرفت ماتریس‌ها *یو کیو کی وی* (معادله (5)). برای تعیین ارتباط بین یک عنصر با سایر عناصر روی دنباله، ضرب نقطه‌ای بین بردار *Q* این عنصر با بردارهای *K* از سایر عناصر محاسبه می‌شود. نتایج، اهمیت نسبی تکه‌ها در دنباله را تعیین می‌کند. سپس نتایج ضرب نقطه‌ای مقیاس‌بندی شده و به یک softmax (معادله (6)) وارد می‌شوند. عملیات مقیاس‌بندی ضرب نقطه‌ای که توسط بلوک SA انجام می‌شود، مشابه ضرب نقطه‌ای استاندارد است، اما آن شامل می‌شود ‎‏ ... بُعد از ‎‏ ... کلید *دی کی* به عنوان الف مقیاس‌بندی عامل. در نهایت، ‎‏ ... ارزش بردار هر پچ جاسازی شده در خروجی softmax ضرب می‌شود تا پچ با ... پیدا شود. توجه زیاد نمرات (معادله (6)). عملیات کامل عبارت است از داده شده توسط اینها معادلات:

[ *س* ، *ک* ، *وی* ] = *زد یو کیو کی وی* ، *یو کیو کی وی* ∈ ر *د* × 3 *د ک* (5)

*الف* = سافت مکس

*کیو کی تی*

√ *دی کی*

، *الف* ∈ ر

*ن* × *ن*

(6)

اس ای ( *ز* ) = *ای* . *وی* (7)

بلوک MSA توجه حاصلضرب نقطه‌ای مقیاس‌بندی‌شده را به‌طور جداگانه برای *h* سر محاسبه می‌کند. با استفاده از عملیات قبلی، اما به جای استفاده از یک مقدار واحد برای پرس‌وجو، کلید و مقدار، از چندین مقدار استفاده می‌شود. نتایج تمام سرهای توجه با هم ترکیب شده و سپس از طریق یک لایه پیش‌خور با وزن‌های قابل یادگیری *W* به بعد مورد نظر تصویر می‌شوند. این عملیات با این معادله بیان می‌شود:

میانگین مربعات عادی ( *z* ) = جمع‌بندی ( SA 1 ( *z* ) ; SA 2 ( *z* ) ؛ . . . SA *h* ( *z* )) *W* ، *W* ∈ آر *اچ دی* کی*​* × *دی* (8)

* + 1. چشم انداز ترانسفورماتور انواع

برای آزمایش تأثیر افزایش اندازه مدل بر دقت طبقه‌بندی، نسخه‌های مختلفی از Vision Transformer در [ [45](#_bookmark60) ] ارائه شده‌اند: "ViT-Base"، "ViT-Large" و "ViT-Huge". این سه نسخه در تعداد لایه‌های رمزگذار متفاوت هستند، ‎‏ ... پنهان بُعد اندازه، ‎‏ ... شماره از توجه سرها استفاده شده توسط ام اس ای لایه، و MLP طبقه بندی کننده اندازه. هر کدام یکی از اینها مدل‌ها است آموزش دیده با الف پچ از اندازه ۱۶ ۱۶ و ۳۲ ۳۲. مدل «ViT-Base» دارای ۱۲ لایه در رمزگذار، با اندازه پنهان ۷۶۸ است و از ۱۲ سر در لایه توجه استفاده می‌کند. نسخه دیگر از اعداد بزرگتری استفاده می‌کند؛ به عنوان مثال، «ViT-Large» دارای ۲۴ لایه، ۱۶ سر توجه و یک بعد پنهان با اندازه ۱۰۲۴ است. «ViT-Huge» دارای ۳۲ لایه، ۱۶ سر توجه و یک اندازه پنهان ۱۲۸۰ است. جدول [۱](#_bookmark3) خلاصه‌ای مقایسه‌ای از نسخه‌های Transformer را نشان می‌دهد.

×

×

**میز ۱.** پارامتر آمار برای ‎‏ ... پایه، بزرگ و عظیم انواع از چشم انداز ترانسفورماتور.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **مدل** | **شماره از لایه‌ها** | **پنهان اندازه دی** | **MLP اندازه** | **سرها** | **شماره از پارامترها** |
| پایه ویتامین | ۱۲ | ۷۶۸ عدد | 3072 | ۱۲ | ۸۶ م |
| ViT- بزرگ | ۲۴ | ۱۰۲۴ عدد | ۴۰۹۶ عدد | ۱۶ | ۳۰۷ م |
| ViT- عظیم | ۳۲ | ۱۲۸۰ | ۵۱۲۰ | ۱۶ | ۶۳۲ م |

نتایج آزمایش روی ترانسفورماتورهای بینایی با اندازه‌های مختلف نشان داده است که استفاده از مدل‌های نسبتاً عمیق‌تر برای دستیابی به دقت بالاتر مهم است. علاوه بر این، انتخاب یک بعد کوچک برای پچ، طول توالی *n را افزایش می‌دهد* که به نوبه خود دقت کلی مدل را بهبود می‌بخشد. یافته مهم دیگر این است که توجه به مراحل اولیه معطوف می‌شود. لایه‌ها از ‎‏ ... چشم انداز ترانسفورماتور می‌تواند شرکت کردن تصویر مناطق در بالا فواصل. این توانایی افزایش می‌یابد به عنوان ‎‏ ... عمق از ‎‏ ... مدل افزایش می یابد. این است متفاوت از ‎‏ ... مبتنی بر CNNها

مدل‌ها، در که زودتر لایه‌ها می‌تواند فقط تشخیص محلی اطلاعات و جهانی اطلاعات فقط در لایه‌های بالاتر شبکه قابل شناسایی هستند. این ویژگی ترانسفورماتور بینایی است حیاتی برای تشخیص ‎‏ ... مربوطه ویژگی‌ها برای طبقه بندی.

* 1. *داده‌ها تقویت استراتژی‌ها*

داده‌افزایی ابزاری ساده اما مؤثر برای افزایش اندازه و تنوع مجموعه داده‌های آموزشی است. این یک گام اساسی برای کارهایی است که دسترسی به یک مجموعه داده بزرگ حاشیه‌نویسی‌شده امکان‌پذیر نیست [ [48](#_bookmark63) ]. داده‌افزایی از تکنیک‌های دستکاری مختلفی برای تولید نمونه‌های آموزشی اضافی از نمونه موجود استفاده می‌کند و در عین حال اعتبار برچسب کلاس اصلی را حفظ می‌کند. آموزش مدل روی داده‌های افزوده به مقابله با مشکل بیش‌برازش کمک می‌کند. و بدین ترتیب بهبود بخشیدن ‎‏ ... استحکام و ‎‏ ... تعمیم توانایی از ‎‏ ... مدل.

تکنیک‌های استاندارد افزایش داده، با اعمال تبدیلات هندسی ساده مانند چرخش، مقیاس‌بندی، برش، تغییر مکان و وارونه کردن، یا استفاده از ترکیبی از آنها، نمونه‌های جدیدی ایجاد می‌کنند. استراتژی‌های تقویت فضای رنگ، مجموعه داده‌ها را با اعمال تبدیلاتی روی فضای رنگ مانند تنظیم روشنایی، کنتراست یا اشباع رنگ تصاویر، گسترش می‌دهند. انتقال سبک عصبی [ [49](#_bookmark64) ] تبدیلات را به شامل شدن ‎‏ ... سطح پایین ویژگی‌ها از ‎‏ ... تصویر، چنین به عنوان بافت. این نقل و انتقالات ‎‏ ... سبک از یک تصویر در ‎‏ ... مجموعه داده‌ها به دیگری تصویر در حالی که نگه داشتن آن معنایی محتوا. یکی رویکرد جالب برای داده‌ها تقویت است ‎‏ ... یکی مبتنی بر روی مولد مدل‌ها، در که مدل‌هایی از این قبیل به عنوان شبکه‌های مولد تخاصمی (GAN) [ [50](#_bookmark65) ] یاد بگیرید ‎‏ ... توزیع از ‎‏ ... داده‌ها به ایجاد کردن مصنوعی نمونه‌ها که هستند تا حد امکان شبیه به تصاویر استخراج شده از مجموعه داده اصلی باشد.

اخیراً تکنیک‌های پیچیده‌تری مبتنی بر پاک کردن تصادفی و ترکیب تصاویر برای تولید نمونه‌های چالش‌برانگیزتر برای مدل معرفی شده‌اند، مانند تکنیک‌های Cutout [ [51](#_bookmark66) ]، Mixup [ [52](#_bookmark67) ] و CutMix [ [53 ].](#_bookmark68) در برش، یک ناحیه تصادفی با اندازه ثابت از تصویر عمداً با پیکسل‌های سیاه یا نویز تصادفی جایگزین می‌شود. این تکنیک ... توسعه یافته به مقابله ‎‏ ... مشکل از مسدود شده اشیاء در صحنه طبقه بندی و تشخیص شیء . الف به طور تصادفی انتخاب شده منطقه است پاک شده به تشویق کردن ‎‏ ... مدل به یاد بگیرید از به جای تکیه بر یک ویژگی بصری خاص، کل بافت تصویر را در نظر می‌گیرد. یکی از مشکلات استفاده از Cutout این است که مسدود کردن می‌تواند بخش مهمی از شیء را پنهان کند و باعث از دست رفتن اطلاعات شود [ [53](#_bookmark68) ]. تکنیک CutMix با برش قسمتی از یک تصویر و جایگزینی آن با قسمتی از تصویر دیگر در مجموعه داده‌ها، بر این مشکل غلبه می‌کند. این می‌تواند از دست رفتن اطلاعات در تکنیک Cutout را کاهش دهد.

با اختلاط، دو تصاویر هستند ادغام شده توسط به صورت خطی درون‌یابی آنها همراه با برچسب‌های کلاس آنها برای ایجاد یک نمونه آموزشی جدید. برای هر دو تقویت CutMix و Mixup، ‎‏ ... زمین حقیقت برچسب است تغییر یافته در مطابق با ‎‏ ... تغییرات اعمال شده به

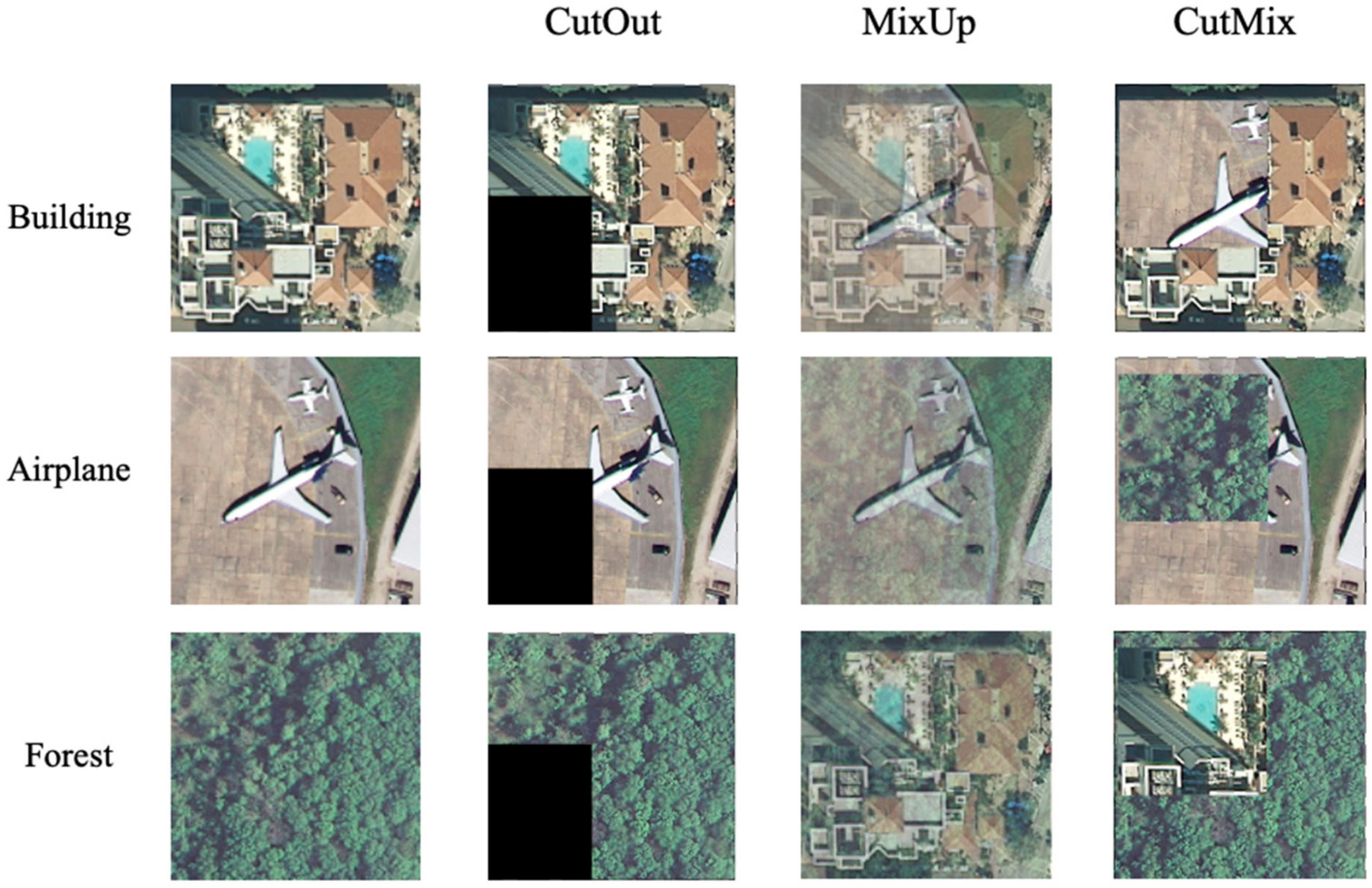
‎‏ ... تصویر. اگر ( *X i* ، *y i* ) و *ایکس جی* ، *وای جی* هستند دو نمونه‌ها کشیده شده به طور تصادفی از ‎‏ ... داده‌های آموزشی و *λ* ∈ [ 0، ۱ ] . (۱) مخلوط کردن تقویت گسترش می‌یابد ‎‏ ... مجموعه داده‌ها توسط درون‌یابی آن دو نمونه‌ها *ایکس من* و *ایکس جی* و آنها مرتبط یک داغ برچسب کدگذاری‌ها *ی من* و *ی ج* با استفاده از ‎‏ ...

دنبال کردن معادلات:

*ایکس* = *λ ایکس آی* + ( 1 − *λ* ) *Xj ( 9* )

*ی* = *λ y i* + ( 1 − *λ* ) *yj ( 10* )

شکل [۳](#_bookmark4) مثال‌هایی از اعمال Cutout، Mixup و CutMix روی مجموعه داده Merced را نشان می‌دهد. انتخاب بهترین استراتژی برای اعمال data augmentation معمولاً یک فرآیند دستی است. روش‌های پیشرفته data augmentation سعی می‌کنند جستجوی تبدیل بهینه برای وظیفه هدف را بدون نیاز به هیچ گونه مداخله انسانی خودکار کنند [ [48](#_bookmark63) ].



**شکل ۳.** مثال‌ها از اعمال کردن داده‌ها تقویت تکنیک‌ها روی مرسد مجموعه داده‌ها.

* 1. *شبکه فشرده سازی*

ترانسفورماتورها معماری عمیق و غنی با میلیون‌ها پارامتر و صدها ... دارند. توجه سرها و چندگانه لایه‌ها. همانطور که می‌تواند باش دیده شده در میز [۱](#_bookmark3) ، ‎‏ ... پایه ویتامین مدل، برای مثال، دارد بیشتر از ۸۰ میلیون پارامترها. در عمومی، مدل‌ها با بزرگ معماری معمولاً نتایج بهتری ارائه می‌دهد. با این حال، پیچیدگی محاسباتی عظیم و عظیم حافظه مورد نیاز مرتبط با اینها مدل‌ها ساختن آنها غیرعملی برای استقرار و مستعد بیش‌برازش.

فشرده‌سازی مدل تکنیک‌ها هدف در تولید کردن الف سبک‌تر نسخه از ‎‏ ... مدل بدون اینکه به دقت اولیه آسیبی برسد. تقطیر دانش و هرس مدل معمولاً مورد استفاده قرار می‌گیرند فشرده سازی رویکردها. با دانش تقطیر، ‎‏ ... اطلاعات کدگذاری شده در یک فرد آموزش دیده مدل معمولاً نامگذاری شده به عنوان ‎‏ ... معلم مدل است منتقل شده به دیگری مدل کوچک‌تر شناخته شده به عنوان ‎‏ ... دانشجو مدل [ [54](#_bookmark69) ]. (این عبارت به نظر می‌رسد ترکیبی از حروف و اعداد باشد و معنای مشخصی ندارد.) دانشجو شبکه تحت نظارت توسط ‎‏ ... شبکه معلم به تدریج یاد می‌گیرد چگونه به تولید کردن نتایج که هستند سازگار با ‎‏ ... نتایج ارائه شده توسط شبکه معلمان. هرس مدل [ [55](#_bookmark70) ] تکنیک دیگری است که برای فشرده‌سازی استفاده می‌شود. سعی می‌کند با حذف پارامترهای اضافی، تعداد پارامترهای مدل را کاهش دهد. یا غیر ضروری اجزا، نگه داشتن فقط ‎‏ ... مهم اجزا. قوطی هرس گرفتن چندین فرم‌ها، چنین به عنوان وزن‌ها کوانتیزه کردن، که استفاده می‌کند کمتر بیت ها به نمایندگی وزن مدل [ [56](#_bookmark71) ]، یا هرس وزن‌ها، که وزن‌های کم‌اطلاعات‌تر را از شبکه حذف می‌کند [ [55](#_bookmark70) ].

مبدل بینایی با معماری افزونه خود با چندین لایه و چندین سر توجه مشخص می‌شود. در این کار، ما یک رویکرد فشرده‌سازی ساده مبتنی بر هرس تدریجی لایه‌های رمزگذار پیشنهاد می‌کنیم. این کار مدل‌های کوچک‌تر با عمق‌های مختلف را از مدل کامل استخراج می‌کند. هدف ما بررسی بده بستان بین عملکرد مدل و عمق مدل است تا فشرده‌ترین معماری که بهترین دقت را ارائه می‌دهد، تعیین شود. در آزمایش‌ها، نشان خواهیم داد که می‌توانیم نیمی از شبکه را هرس کنیم و در عین حال دقت طبقه‌بندی رقابتی را حفظ کنیم.

در الگوریتم ۱ زیر، مراحل اصلی آموزش Vision Transformer را ارائه می‌دهیم:

**الگوریتم ۱:** چشم انداز ترانسفورماتور

Input: Training images: {*Xi*, *yi*}

*ن من* = ۱

خروجی: پیش‌بینی شده برچسب‌ها از ‎‏ ... آزمون تنظیم شده.

1. *اندازه دسته‌ها را روی ۱۰۰* تنظیم کن ، بهینه‌ساز Adam (نرخ یادگیری: ۰.۰۰۰۳)، تعداد تکرارها را روی ۳۰ تنظیم کن، تصویر ابعاد ۲۲۴ *یا* ۳۸۴.
2. تنظیم ‎‏ ... شماره از مینی دسته‌ها به عنوان: *ن ب* = *n* / *اندازه دسته‌ای*
3. برای تکرار = ۱: شماره از تکرارها
   1. برای دسته‌ای = ۱ : *ن ب*

انتخاب الف دسته‌ای از ‎‏ ... آموزش تنظیم شده،

•

تولید کردن دیگری دسته‌ای از تقویت شده تصاویر با استفاده از الف خاص تقویت روش، قطار ‎‏ ... مدل روی ‎‏ ... اصلی و تقویت شده تصاویر توسط کوچک سازی ‎‏ ... آنتروپی متقاطع زیان.

•

•

* + - انتشار معکوس ‎‏ ... زیان.
    - به‌روزرسانی ‎‏ ... مدل پارامترها.

1. طبقه بندی آزمون تصاویر

# تجربی نتایج

* 1. *مجموعه داده‌ها توضیحات*

در ما آزمایش‌ها، سه شناخته شده سنجش از دور مجموعه داده‌ها هستند استفاده شده برای ارزیابی: مجموعه داده‌های کاربری زمین Merced [ [57](#_bookmark72) ]، مجموعه داده‌های تصویر هوایی (AID) [ [58](#_bookmark73) ]، و مجموعه داده‌های Optimal-31 [ [41](#_bookmark56) ]. ویژگی‌های این سه مجموعه داده در جدول [2 فهرست شده است و نمونه‌هایی از هر مجموعه داده در شکل](#_bookmark6) [4](#_bookmark7) نشان داده شده است .

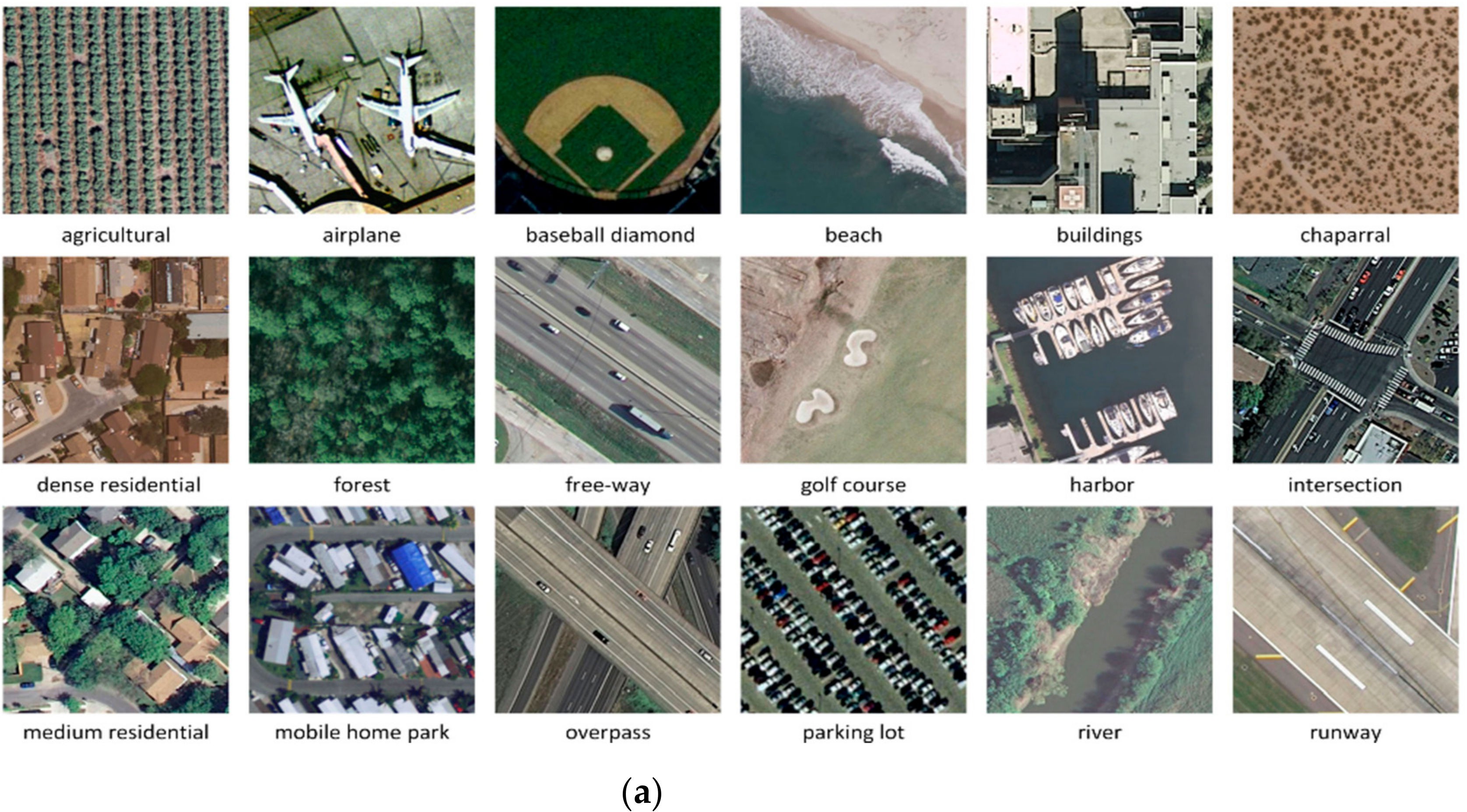
**میز ۲.** مشخصه از ‎‏ ... مجموعه داده‌ها.

**شماره مجموعه داده از شماره کلاس‌ها از تصاویر به ازای هر تصویر کلاس اندازه سال**

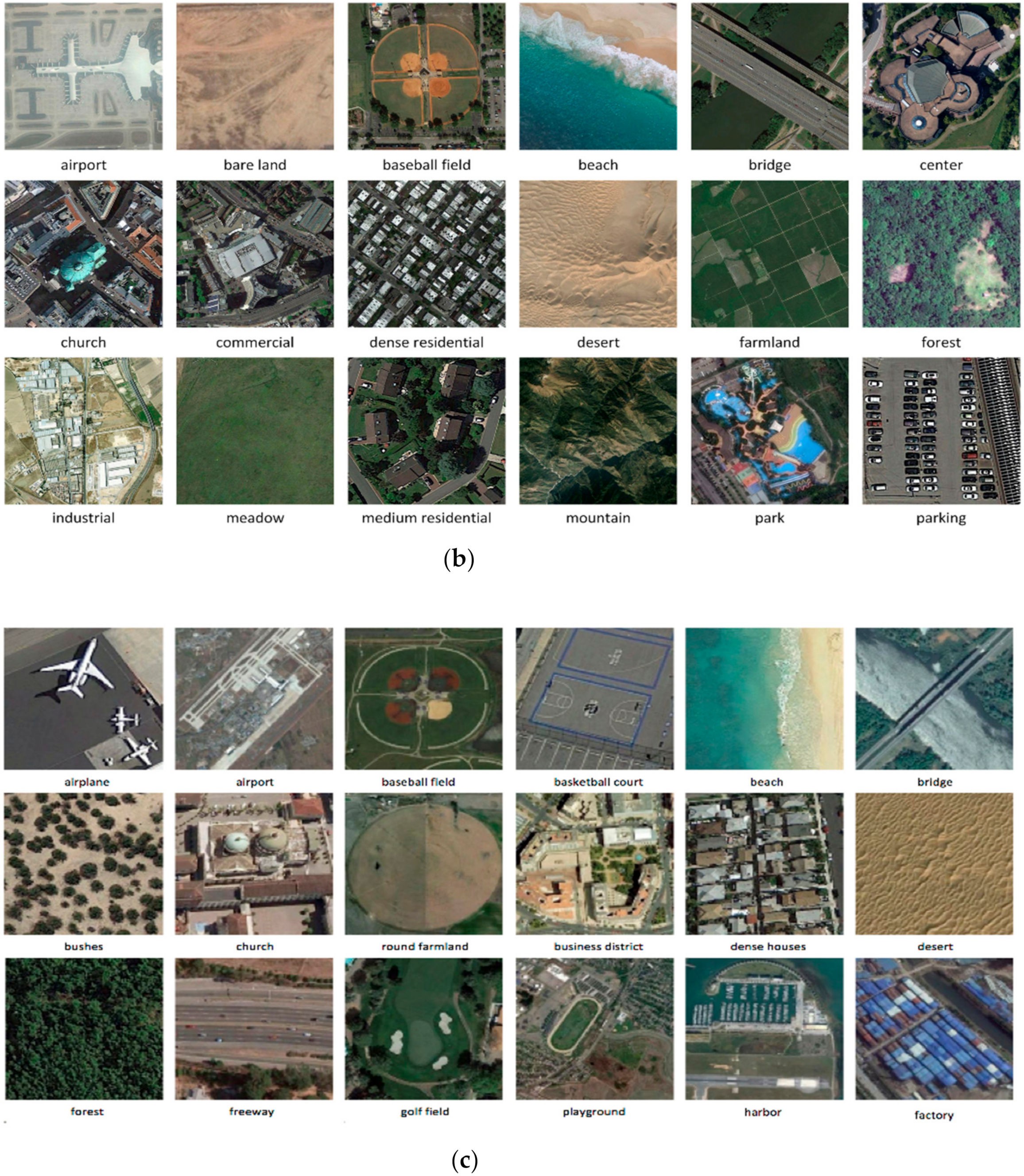
مرسد ۲۱ ۱۰۰ ۲۵۶ × ۲۵۶ ۲۰۱۰

AID 30 220~420 600 × ۶۰۰ ۲۰۱۷

بهینه ۳۱ ۳۱ ۶۰ ۲۵۶ × ۲۵۶ ۲۰۱۹



**شکل ۴.** *ادامه*​



**شکل ۴.** برخی مثال تصاویر از ( **الف** ) مرسد مجموعه داده‌ها. ( **ب** ) اید مجموعه داده‌ها. ( **ج** ) بهینه- ۳۱.

مجموعه داده Merced: این مجموعه داده در سال ۲۰۱۰ منتشر شد و شامل ۲۱۰۰ تصویر RGB از ۲۱ کاربری زمین صحنه کلاس‌ها هر کدام کلاس متشکل از از ۱۰۰ تصاویر از اندازه ۲۵۶ ۲۵۶ پیکسل‌ها با

×

وضوح ۰.۳ متر تصاویر از نقشه ملی سازمان زمین‌شناسی ایالات متحده استخراج شده‌اند.

هوایی تصویر مجموعه داده‌ها (ایدز) مجموعه داده‌ها: اید مجموعه داده‌ها است الف در مقیاس بزرگ مجموعه داده‌ها از ۱۰،۰۰۰ برای تصاویر هوایی منتشر شده در سال ۲۰۱۷ توسط دانشگاه ووهان. مجموعه داده‌ها شامل 30 مورد است متفاوت کلاس‌ها از ۲۲۰ به ۴۲۰ تصاویر به ازای هر کلاس تصاویر بودند برش خورده از گوگل ارث تصویرسازی اندازه‌گیری ۶۰۰ ۶۰۰ پیکسل‌ها با الف وضوح تصویر متغیر از ۸ متر به درباره

×

۰.۵ م.

مجموعه داده بهینه-۳۱: این مجموعه داده از تصاویر گوگل ارث گرفته شده است که پوشش ... ۳۱ کلاس صحنه. هر کلاس شامل ۶۰ تصویر با اندازه ۲۵۶ است ۲۵۶ پیکسل در فضای رنگی RGB. وضوح پیکسلی تصاویر 0.3 متر است.

×

* 1. *تجربی راه‌اندازی*

ما انجام شده سه مجموعه‌ها از آزمایش‌ها. در ‎‏ ... اول تنظیم شده، ما استفاده شده متفاوت داده‌ها استراتژی‌های تقویت برای ارزیابی عملکرد Vision Transformer با داده‌های تقویت‌شده. لازم به یادآوری است که ما از تلفات آنتروپی متقاطع استاندارد برای یادگیری وزن‌های شبکه استفاده کردیم. در آزمایش دوم، تعداد لایه‌های رمزگذار را تغییر دادیم و رابطه بین عمق شبکه و عملکرد مدل را بررسی کردیم. سپس، تأثیر تغییر اندازه تصویر بر دقت کلی مدل را بررسی کردیم. در نهایت، ما مقایسه شده ما نتایج علیه چندین مدرن روش‌ها.

در تمام آزمایش‌ها، ما مدل ViT-Base را با تنظیمات زیر از [ [45](#_bookmark60) ] اتخاذ کردیم. این مدل شامل 12 لایه رمزگذار است که هر کدام دارای 12 سر توجه هستند. دارای بُعد تعبیه ۷۶۸ و زیرشبکه پیش‌خور با اندازه ۳۰۷۲ است. ما از مدلی که از قبل روی Imagenet-21k آموزش دیده و سپس روی Imagenet-1k تنظیم دقیق شده بود، استفاده کردیم. برای تنظیم دقیق آن روی سنجش از دور صحنه داده‌ها، ما آموزش دیده آن برای ۳۰ تکرارها و استفاده شده الف مینی‌بچ اندازه از ۱۰۰. (یا: ۱۰۰) ما

آن را با روش آدام بهینه کردیم و نرخ یادگیری را روی ۰.۰۰۰۳ تنظیم کردیم. در ابتدا اندازه تصویر را روی ۲۲۴ ثابت کردیم. ۲۲۴ و اندازه پچ به ۱۶ ۱۶ و دنباله‌ای با طول ۱۹۶ توکن به دست آمد.

× ×

برای مقایسه اهداف، ما ارزیابی شده ‎‏ ... عملکرد از ‎‏ ... روش در اصطلاحات از دقت کلی استاندارد (OA)، که نشان دهنده تعداد تصاویر به درستی طبقه بندی شده نسبت به تعداد کل تصاویر است.

ما تمام آزمایش‌ها را روی یک HP Omen Station با مشخصات زیر انجام دادیم: مرکزی پردازش واحد (پردازنده مرکزی) اینتل هسته (™) i9-7920 پردازنده @ ۲.۹ گیگاهرتز با رم ۳۲ گیگابایتی و واحد پردازش گرافیکی (GPU) NVIDIA GeForce GTX 1080 Ti (با ۱۱ گیگابایت حافظه GDDR5X). تمام کدها با استفاده از Pytorch، که یک کتابخانه شبکه عصبی عمیق متن‌باز نوشته شده با پایتون است، پیاده‌سازی شدند.

×

* 1. *آزمایش ۱: مقدماتی تحلیل*

برای تحلیل اولیه‌ی Vision Transformer، ما یک رژیم آموزشی سطح پایین را دنبال کردیم و آزمایش را با حداقل داده‌ها انجام دادیم. به طور خاص، برای مجموعه داده‌های AID، موارد زیر را انتخاب کردیم: درباره ۳۳ نمونه‌ها از هر کدام کلاس برای آموزش، که متشکل از ۱۰٪ از ‎‏ ... برای هر دو مجموعه داده Merced و Optimal31، از هر کلاس 30 نمونه استخراج کردیم که به ترتیب 30٪ و 50٪ از مجموعه داده‌های اول و دوم را تشکیل می‌دادند.

ما آموزش دیده ‎‏ ... شبکه روی ‎‏ ... اصلی و تقویت شده تصاویر برای ۳۰ تکرارها جدول [۳](#_bookmark8) دقت طبقه‌بندی حاصل از اعمال تکنیک‌های مختلف افزایش داده را نشان می‌دهد. افزایش داده استاندارد از چرخش، وارونه کردن عمودی و افقی و تنظیم تصادفی روشنایی و رنگ تصویر استفاده می‌کند. برای برش در این تکنیک، تعداد سوراخ‌ها را روی هشت و اندازه ناحیه برش را روی ۱۰ تنظیم می‌کنیم. 10 پیکسل. برای CutMix، نسبت اختلاط از توزیع یکنواخت [0،1] نمونه‌برداری شد. در نهایت، داده‌افزایی ترکیبی به طور تصادفی یکی از سه تکنیک داده‌افزایی (استاندارد، کات میکس و برش) برای هر کدام دسته‌ای در طول ‎‏ ... آموزش فاز

×

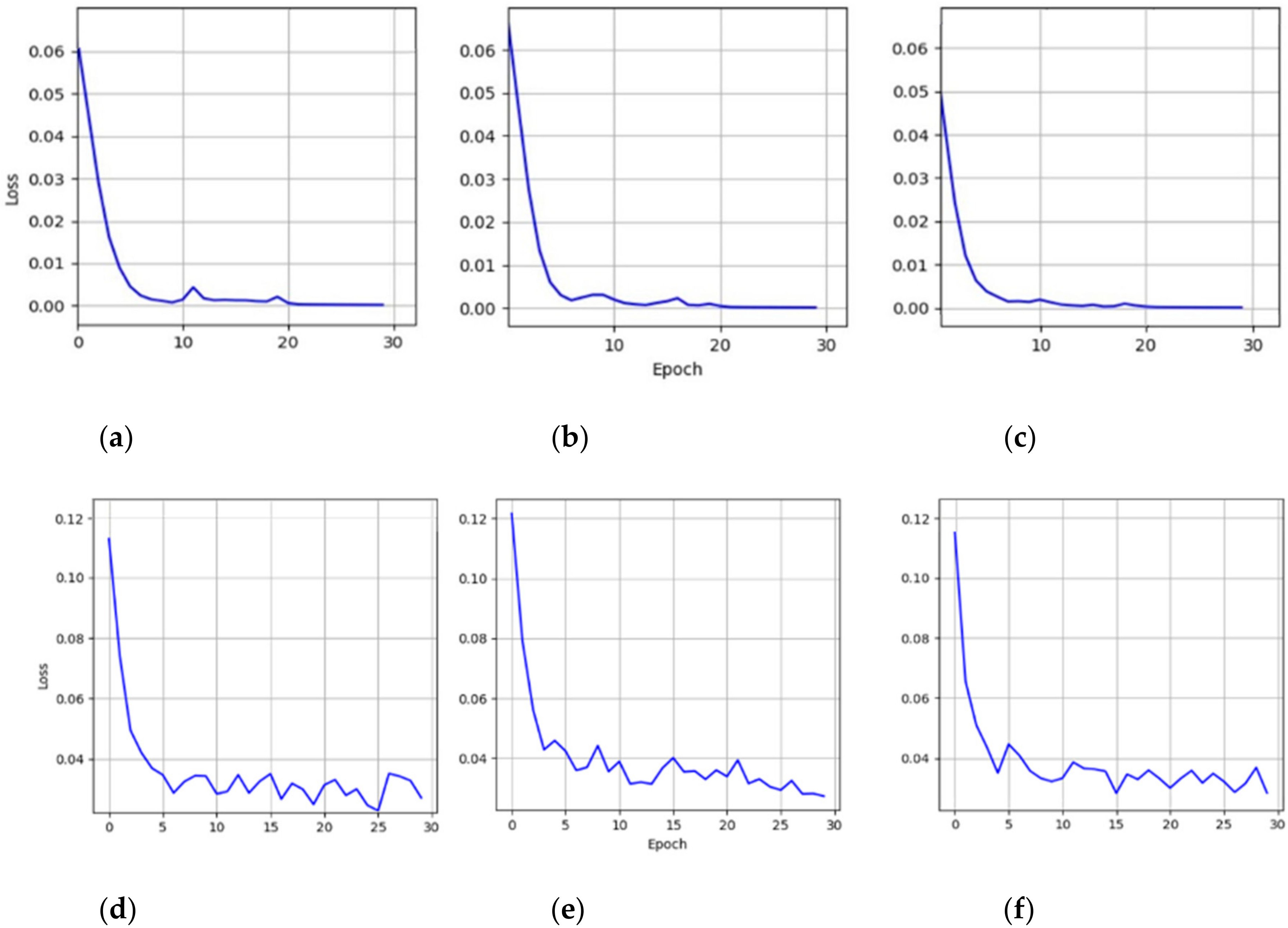
**میز ۳.** طبقه‌بندی نتایج در: (الف): مرسد مجموعه داده‌ها (30٪ قطار مجموعه)، اید (۱۰٪) قطار) و بهینه ۳۱ (۵۰٪ مجموعه آموزشی). اندازه تصویر: ۲۲۴ × ۲۲۴.

**با تقویت**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **مجموعه داده‌ها** | **تمیز تصاویر** | **استاندارد** | **کات میکس** | **برش** | **هیبریدی** |
| مرسد | ۹۴.۵۵ | ۹۶.۳۲ | ۹۶.۶۶ | ۹۵.۴۴ | ۹۶.۷۳ |
| اید | ۸۹.۳۱ | ۹۲.۰۶ | ۹۰.۵۰ | ۹۱.۶۲ | ۹۱.۷۶ |
| بهینه31 | ۸۸.۲۷ | ۹۱.۴۳ | ۹۲.۴۴ | ۹۲.۲۵ | ۹۲.۹۷ |
| میانگین | ۹۰.۷۱ | ۹۳.۲۷ | ۹۳.۲۰ | ۹۳.۳۰ | ۹۳.۸۲ |

نتایج در جدول [3](#_bookmark8) به وضوح اثربخشی استفاده از افزایش داده را همانطور که در ادبیات به طور گسترده شناخته شده است، نشان می‌دهد. به طور کلی، همه استراتژی‌ها نتایج نزدیکی ارائه دادند، اما برای مجموعه داده‌های Merced و Optimal31، استفاده از تقویت داده ترکیبی کمی بهتر عمل کرد. نتایج با دقت از ۹۶.۷۳٪ و ۹۲.۹۷٪، به ترتیب. استاندارد افزایش با ۹۲.۰۶٪ عملکرد کمی بهتر از سایر تکنیک‌ها برای مجموعه داده AID داشت. معمولاً، و به عنوان ‎‏ ... نتایج از همه ‎‏ ... سه مجموعه داده‌ها پیشنهاد دادن، با استفاده از الف ترکیب از داده‌ها استراتژی‌های تقویت ، تقریباً بهترین رفتار را ارائه می‌دهند.

شکل [۵](#_bookmark9) تکامل تابع زیان را در طول آموزش با و بدون [مسیر] نشان می‌دهد. افزایش داده. می‌توان مشاهده کرد که آموزش مدل روی تصاویر اصلی باعث از دست رفتن همگرا شدن صاف تر و سریعتر. در تضاد، چه زمانی ‎‏ ... مدل است آموزش دیده روی ‎‏ ... در تصاویر افزوده، میزان تلفات پس از تکرارهای گرم شدن نوسان می‌کند و همگرایی آن زمان بیشتری می‌برد.

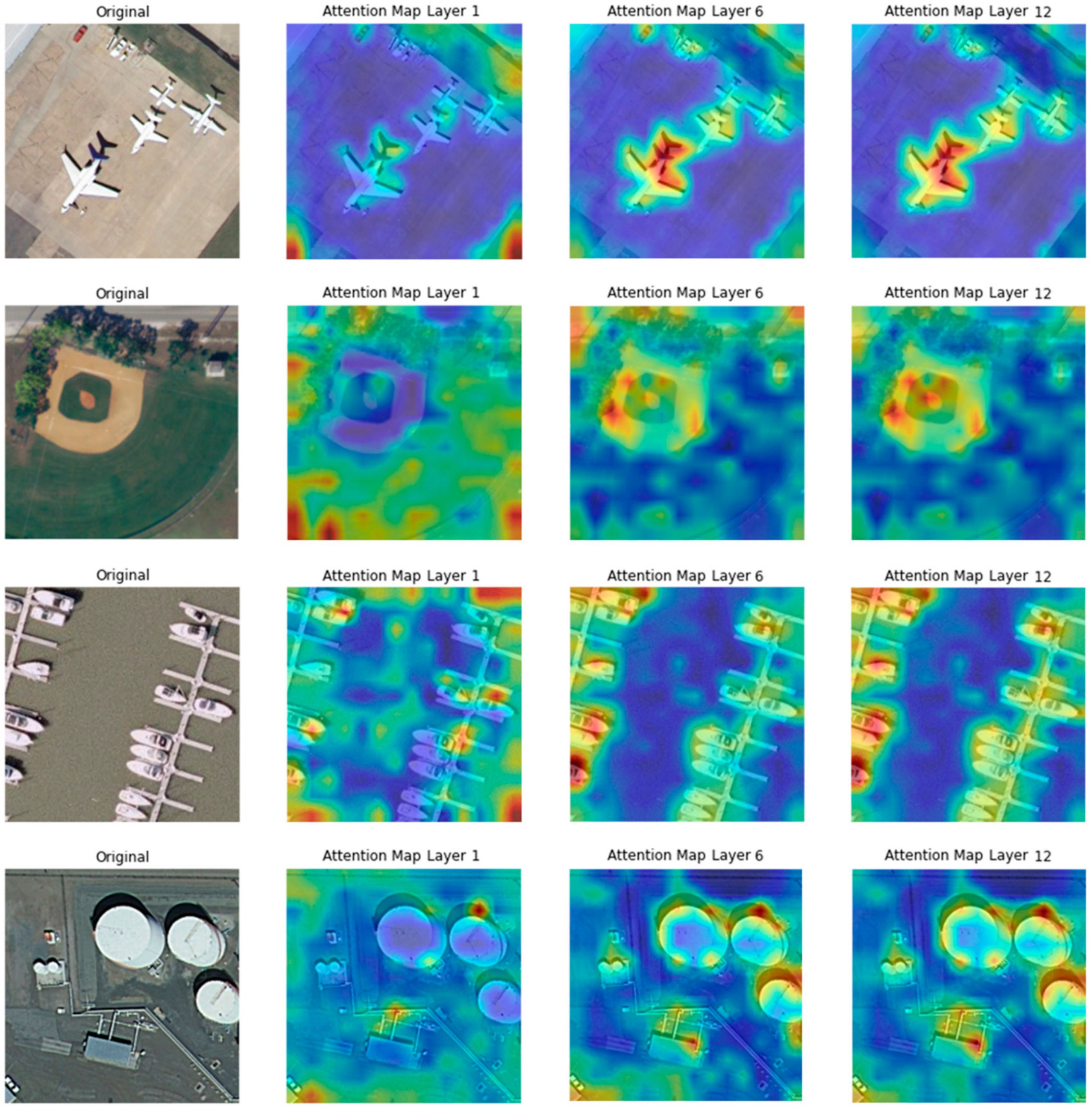


**شکل ۵.** تابع زیان بدون افزایش داده برای: ( **الف)** Merced، ( **ب** ) AID، و ( **ج** ) Optimal31، و با افزایش داده برای: ( **د** ) مرسد، ( **ه** ) AID، و ( **و** ) Optimal31.

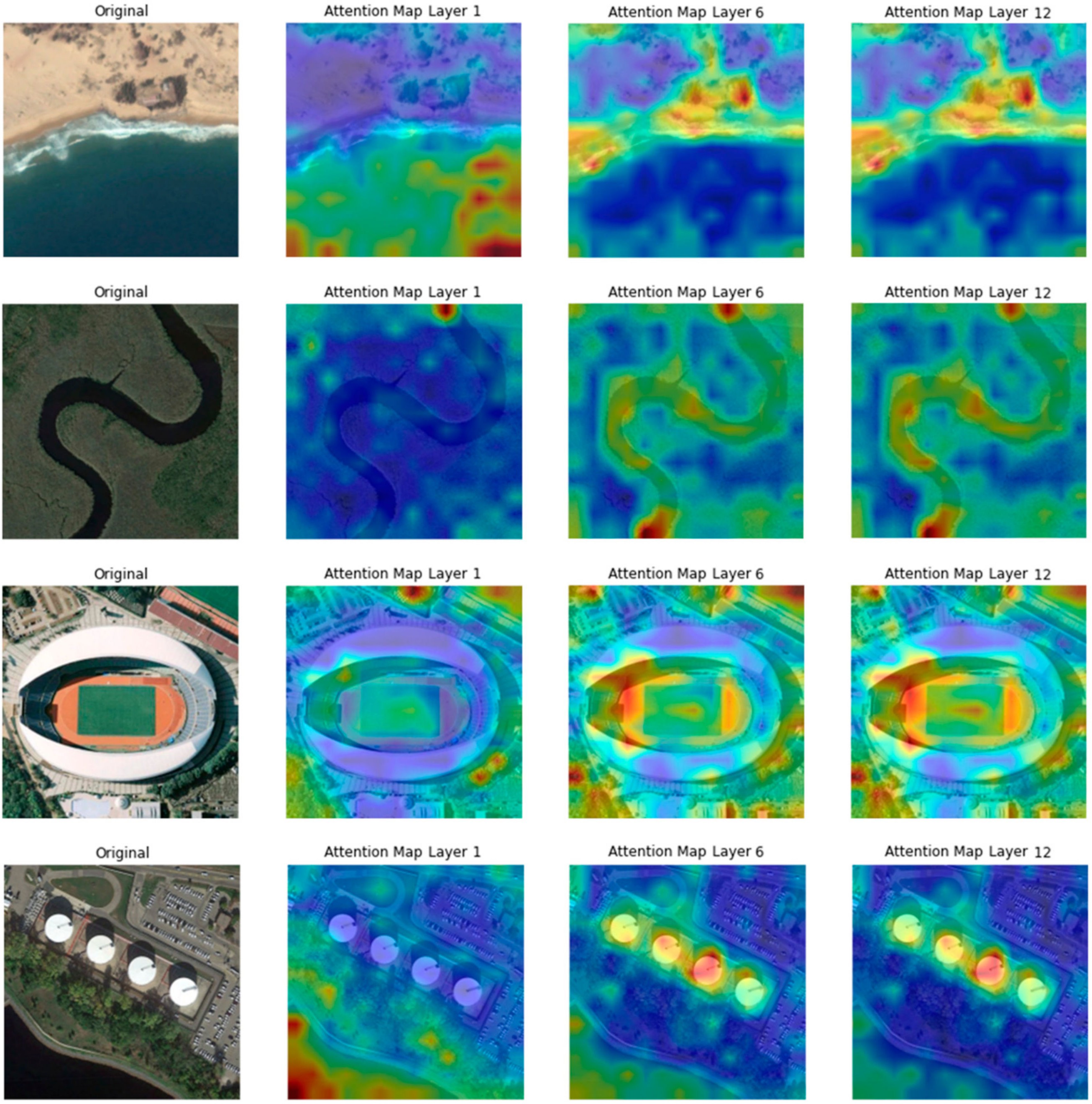
* 1. *آزمایش ۲: شبکه فشرده سازی*

در مجموعه دوم آزمایش‌ها، نقش هر لایه در رمزگذار را بیشتر تجزیه و تحلیل کردیم. ابتدا، مدل را با حداکثر تعداد لایه‌ها (یعنی ۱۲ لایه) آموزش دادیم. سپس، آزمایش‌ها را با همان پارامترها تکرار کردیم، با این تفاوت که خروجی هر لایه میانی را گرفتیم و آن را مستقیماً به طبقه‌بندی‌کننده منتقل کردیم و لایه‌های بالایی را کنار گذاشتیم. برای درک بهتر رفتار شبکه و ناحیه‌ای که سرهای توجه در هر لایه در آن قرار دارند، نمایش‌های خروجی را استخراج کردیم و نقشه‌های توجه هر لایه را برای مجموعه داده‌های Merced و AID در شکل‌های [۶ تجسم کردیم.](#_bookmark10) و به ترتیب [۷](#_bookmark11) .

شکل [۶](#_bookmark10) چهار نمونه از مجموعه داده Merced را به همراه خروجی‌های لایه‌های اول، ششم و دوازدهم نشان می‌دهد. می‌توانیم ببینیم که شبکه به تدریج یاد می‌گیرد روی مناطقی تمرکز کند که بیشترین اطلاعات نماینده از کلاس را دارند. به عنوان مثال، در فرودگاه نمونه ‎‏ ... شبکه نشان می‌دهد برخی توجه روی هواپیماها در لایه ۱. این به تدریج بهبود یافته در ‎‏ ... متعاقباً لایه‌ها برای مثال، برای ‎‏ ... بیسبال کلاس، ‎‏ ... شبکه در لایه اول، بیشتر بر اطلاعات نامرتبط تمرکز شد و سپس مدل تلاش کرد تا مناطق متمایز مربوط به کلاس بیسبال را ثبت کند. برای کلاس بندر، با افزایش عمق رمزگذار، مدل تمایل داشت تمرکز بیشتری بر ناحیه قایق‌ها داشته باشد. این تغییر در تمرکز از لایه ۱ به لایه دیگر به شدت قابل مشاهده بود. ۶. با این حال، پس از لایه ۶ ما می‌تواند دیدن که ‎‏ ... توجه به غیرمرتبط مناطق بود کاهش یافته است. این موضوع در بخش بعدی بیشتر تأیید خواهد شد.



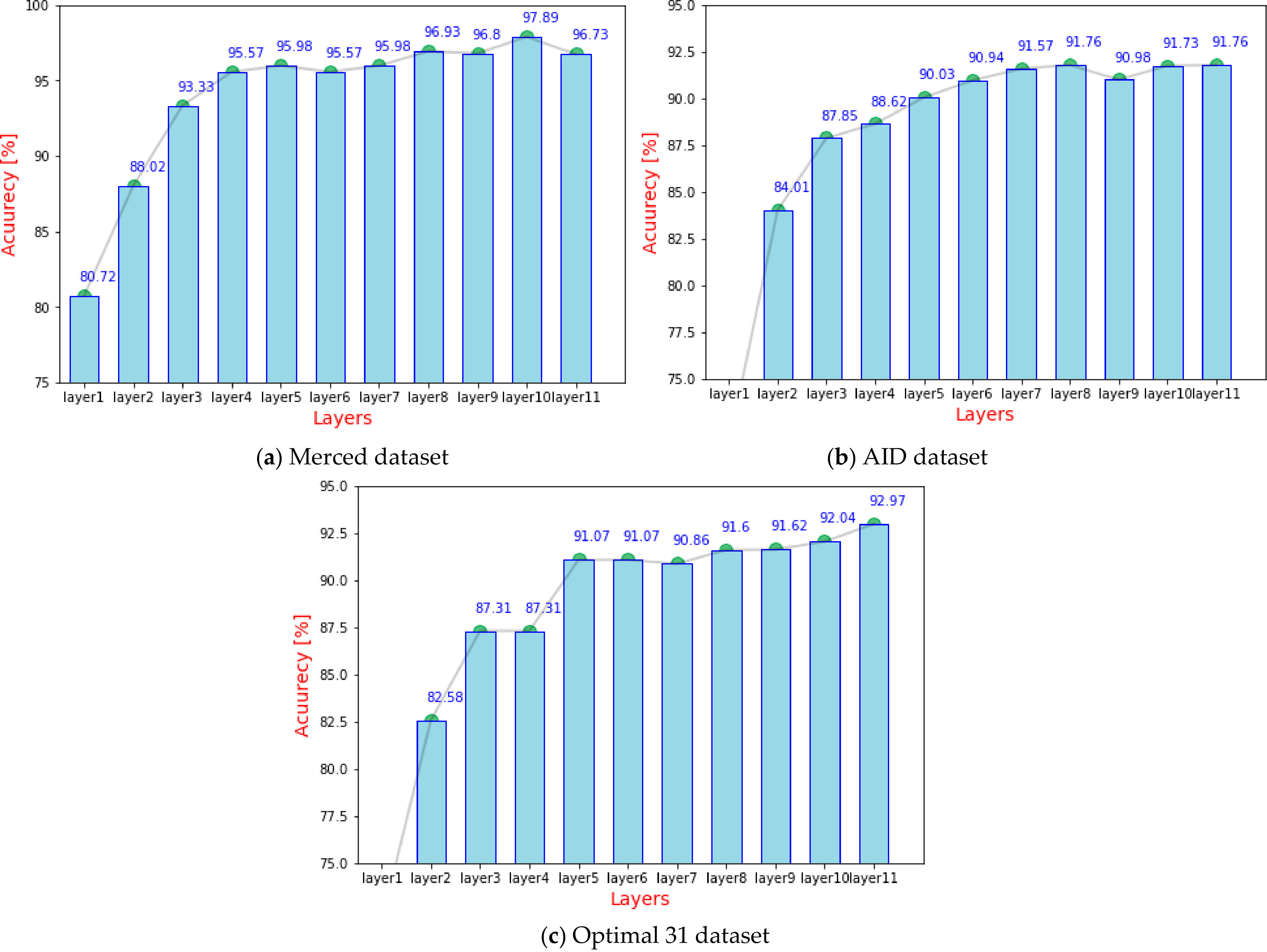
**شکل ۶.**  کلاس توجه نقشه‌ها نتیجه داد از متفاوت رمزگذار لایه‌ها برای ‎‏ ... مرسد مجموعه داده‌ها.



**شکل ۷.**  کلاس توجه نقشه‌ها نتیجه داد از متفاوت رمزگذار لایه‌ها برای ‎‏ ... اید مجموعه داده‌ها.

شکل [۷](#_bookmark11) چهار تصویر از مجموعه داده AID را به همراه خروجی سه لایه رمزگذار مختلف (لایه ۱، ۶ و لایه آخر) نشان می‌دهد. همانطور که مشاهده می‌شود، برای کلاس ساحل، شبکه در لایه ۱ عمدتاً تمرکزها روی ‎‏ ... دریا مناطق. سپس، ‎‏ ... بعدی رمزگذار لایه‌ها یاد بگیرید که به تدریج توجه را به خط ساحلی معطوف کنید و در عین حال به تدریج مناطق نامربوط را نادیده بگیرید. در علاوه بر این، ما مشاهده شده که ‎‏ ... توجه نقشه‌ها ارائه شده توسط لایه ۶ هستند از نظر بصری مشابه به ‎‏ ... یکی ارائه شده در ‎‏ ... آخرین لایه ما مشاهده شده همچنین الف مشابه رفتار برای کلاس رودخانه که در آن شبکه بر روی ناحیه رودخانه در لایه ۶ تمرکز دارد و توجه کمی به آن معطوف است بهبود می‌بخشد در ‎‏ ... آخرین لایه برای ‎‏ ... ورزشگاه تصویر، به عنوان ‎‏ ... رمزگذار می‌شود هرچه عمیق‌تر شود، یاد می‌گیرد بخش‌های متمایزکننده‌ای را که مربوط به کلاس استادیوم هستند، محلی‌سازی کند. در نهایت، برای کلاس تانک، مشاهده کردیم که شبکه روی اشیاء نامرتبط تمرکز می‌کند. در ‎‏ ... اول لایه اما بود قادر به کنسانتره روی ‎‏ ... مخزن اشیاء در لایه ۶. این یعنی که با عمیق‌تر شدن رمزگذار، توجه بهبود می‌یابد.

از الف کمی نقطه از مشاهده، شکل [۸](#_bookmark12) نشان می‌دهد ‎‏ ... طبقه بندی دقت ها در هر لایه از رمزگذار برای مجموعه داده‌های Merced، AID و Optimal 31 به دست آمده است. به طور کلی، ما می‌توان دید که رمزگذارهای عمیق معمولاً عملکرد بهتری دارند و عملکرد طبقه‌بندی با افزایش تعداد لایه‌ها به طور مداوم افزایش می‌یابد. شکل نشان می‌دهد که استفاده از رمزگذارهایی با حداقل ۵ لایه برای رسیدن به دقت طبقه‌بندی ۹۰٪ در تمام مجموعه داده‌ها کافی است. لایه‌های بعدی از ۶ تا ۱۲، دقت را ۲٪ بهبود می‌بخشند. این نشان می‌دهد که لایه‌های اولیه در مدل Vision Transformer نقش کلیدی در استخراج نمایش متمایزکننده مورد نیاز برای طبقه‌بندی دارند.



**شکل ۸.** نسبی تغییر در مدل طبقه بندی دقت با احترام به ‎‏ ... رمزگذار لایه‌ها برای ‎‏ ... ( **الف** ) مرسد، ( **ب** ) AID، و ( **ج** ) مجموعه داده‌های Optimal-31.

میانگین نتایج سه مجموعه داده نشان می‌دهد که هرس مدل تا لایه ۱۰ بهترین عملکرد را با دقت متوسط ۹۳.۸۸٪ در مقایسه با ۹۳.۸۲٪ با مدل کامل ارائه می‌دهد. بنابراین، برای طبقه‌بندی صحنه، آخرین لایه مدل Vision Transformer را می‌توان بدون تأثیر بر عملکرد مدل حذف کرد.

به طور خاص، برای مجموعه داده Optimal31، بهترین دقت طبقه‌بندی را می‌توان از آخرین لایه با دقت ۹۲.۹۷٪ بدست آورد. با این حال، جالب است که مشاهده کنیم که بالاترین دقت را می‌توان از لایه‌های قبلی برای دو مجموعه داده دیگر بدست آورد. به عنوان مثال، یک رمزگذار با ۱۰ لایه، بهترین دقت طبقه‌بندی را برای مجموعه داده Merced با دقت ۹۷.۸۹٪ ارائه می‌دهد. برای مجموعه داده AID، لایه ۸ و ۱۲ به طور یکسان با ۹۱.۷۶٪ بهترین نتایج را ارائه می‌دهند. این نتایج با نتایج کیفی بدست آمده از نقشه‌های توجه سازگار است. در بخش بعدی، نشان خواهیم داد که استفاده تنها از 50٪ لایه‌ها می‌تواند دقت طبقه‌بندی رقابتی را به همراه دارد.

# بحث

ما همچنین تأثیر تغییر اندازه تصویر بر عملکرد را بررسی می‌کنیم. مدل. به این پایان، ما تکرار ‎‏ ... آزمایش‌ها با استفاده از تصاویر با دو متفاوت اندازه‌ها،

۲۲۴ ۲۲۴ و ۳۸۴ ۳۸۴. در واقع، مدل‌های Vision Transformer روی مجموعه داده ImageNet با اندازه تصویر ۳۸۴ از قبل آموزش داده شده بودند. ۳۸۴.

×

× ×

دقت کلی و زمان اجرای آزمایش‌ها در جدول خلاصه شده است [۴](#_bookmark14) . نتایج به وضوح نشان دادن یک افزایش در تصویر اندازه چه زمانی ‎‏ ... مدل است آموزش دیده با بزرگ تصویر اندازه. با این حال، افزایش ‎‏ ... اندازه می‌تواند به طور قابل توجهی بالا بردن ‎‏ ... آموزش زمان. به طور متوسط، با استفاده از بزرگتر تصاویر دارد بهبود یافته ‎‏ ... نتیجه توسط ۰.۹۳٪ اما دو برابر شده ‎‏ ... زمان آموزش از ۳۲ به ۶۷ دقیقه برای ‎‏ ... بهینه31 مجموعه داده‌ها، این هزینه دارد الف خفیف بهبود با دقتی تنها 0.03٪.

**میز ۴.** دقت‌های کلی و زمان‌های آموزش به‌دست‌آمده با استفاده از اندازه‌های مختلف تصویر روی ریموت‌های مختلف مجموعه داده‌ها.

|  |  |
| --- | --- |
| **مجموعه داده ۲۲۴ *×* ۲۲۴** | **۳۸۴ *×* ۳۸۴** |
| مرسد ۹۶.۷۳ | ۹۷.۴۳ |
| ۳۰ دقیقه | ۴۶ دقیقه |
| AID 91.76 | ۹۲.۹۴ |
| ۴۳ دقیقه | ۸۸ دقیقه |
| بهینه ۳۱ ۹۲.۷۶ | ۹۲.۷۹ |
| ۲۵ دقیقه | ۶۸ دقیقه |
| میانگین ۹۳.۴۵ | ۹۴.۳۸ |
| ۳۲.۶۶ دقیقه | ۶۷.۳۳ دقیقه |

در نهایت، ما مقایسه ‎‏ ... نتایج از ما روش با ‎‏ ... مدرن نتایج گزارش شده است دور در ‎‏ ... ادبیات. اینها روش‌ها هستند ‎‏ ... توجه مکرر کانولوشن شبکه (ARCNet) [ [41](#_bookmark56) ]، که در آن چندین ویژگی توجه با استفاده از معماری CNN-LSTM تولید می‌شوند. GoogleNet ویژگی‌های طبقه‌بندی شده با طبقه‌بندی کننده SVM را استخراج کرد [ [58](#_bookmark73) ]. دو طرفه دروازه‌دار شبکه که استفاده می‌کند سلسله مراتبی ویژگی تجمیع (جی‌بی نت) [ [31](#_bookmark46) ]. ادغام کوواریانس چندلایه انباشته‌شده (MSCP) [ [59](#_bookmark74) ]، که در آن ویژگی‌های لایه‌های مختلف CNN از پیش آموزش‌دیده با استفاده از ادغام کوواریانس ترکیب شده و با استفاده از SVM طبقه‌بندی می‌شوند. علاوه بر این، ما اضافه کردن ‎‏ ... نتایج از تنظیم دقیق VGG16 و گوگل نت مدل‌ها [ [60](#_bookmark75) ] و مدل‌هایی که با یک طبقه‌بندی‌کننده کمکی تنظیم دقیق شده‌اند [ [30](#_bookmark45) ].

جدول [۵](#_bookmark15) مقایسه‌های دقیقی را به ترتیب برای مجموعه داده‌های Merced، AID و Optimal-31 نشان می‌دهد. علاوه بر این سه مجموعه داده، نتایج خود را با مجموعه داده شناخته شده NWPU که از ۴۵ کلاس شامل ۳۱۵۰۰ تصویر سنجش از دور تشکیل شده است، مقایسه می‌کنیم. بسته به تقسیم‌بندی داده‌های گزارش شده در مقالات، تقسیم‌بندی آموزش-آزمایش را به طور متفاوتی تنظیم می‌کنیم. برای هر کدام مجموعه داده‌ها. ما اصطلاح شده ‎‏ ... پیشنهادی روش به عنوان V16\_21k (224 ۲۲۴)، و V16\_21k (384) ۳۸۴) برای Vision Transformer که تصاویر را به ۱۶ قسمت تقسیم می‌کند ۱۶، از پیش آموزش دیده روی مجموعه داده Imagenet-21k و تنظیم دقیق با تصاویر با اندازه ۲۲۴ ۲۲۴ و ۳۸۴ به ترتیب ۳۸۴. نتایج در جدول [5](#_bookmark15) نشان می‌دهد که شبکه نتایج جالبی برای

× ×

× ×

×

همه مجموعه داده‌ها. در خاص، ‎‏ ... پیکربندی با بزرگ تصویر اندازه و کوچکتر پچ اندازه، عملکرد برتر را به دست می‌دهد. از نظر زمان محاسبه، شبکه برای مرسد به 153 زمان نیاز دارد. دقیقه؛ کمک: ۳۴۷ دقیقه؛ بهینه ۳۱: ۲۲۰ دقیقه؛ و NWPU: ۴۶۵ دقیقه علاوه بر این، جدول [5](#_bookmark15) نشان می‌دهد که شبکه پس از هرس کردن ۵۰٪ از لایه‌های خود، نتایج بسیار رقابتی ارائه می‌دهد.

**میز ۵.** مقایسه با مدرن روش‌ها.

**روش مرسد**

**مجموعه داده‌ها**

**AID Optimal31**

**NWPU**

**(50٪ قطار)**

**(20٪) قطار)**

**(۸۰٪) قطار)**

**(۱۰٪) قطار)**

ARCNet-VGG16 [ [41](#_bookmark56) ] 96.81 ± ۰.۱۴ ۸۸.۷۵ ± ۰.۴۰ ۹۲.۷۰ ± ۰.۳۵ -

آرک‌نت- الکس نت [ [41](#_bookmark56) ] - - 85.75 ± ۰.۳۵ -

آرک‌نت- ریس‌نت [ [41](#_bookmark56) ] - - 91.28 ± ۰.۴۵ -

گوگل‌لی‌نت+ماشین بردار پشتیبان [ [58](#_bookmark73) ] 92.70 ± ۰.۶۰ ۸۳.۴۴ ± ۰.۴۰ - -

جی‌بی نت + جهانی ویژگی [ [31](#_bookmark46) ] 97.05 ± ۰.۱۹ ۹۲.۲۰ ± ۰.۲۳ ۹۳.۲۸ ± ۰.۲۷ -

VGG-16+MSCP [ [59](#_bookmark74) ] 98.36 ± ۰.۵۸ ۹۱.۵۲ ± ۰.۲۱ - -

تنظیم دقیق VGG16 [ [31](#_bookmark46) ] 96.57 ± ۰.۳۸ ۸۹.۴۹ ± ۰.۳۴ ۸۹.۵۲ ± ۰.۲۶ ۸۷.۱۵ ± ۰.۴۵

تنظیم دقیق گوگل‌لی‌نت [ [60](#_bookmark75) ] - - 82.57 ± ۰.۱۲ ۸۲.۵۷ ± ۰.۱۲

Inception-v3-aux [ [30](#_bookmark45) ] 97.63 ± ۰.۲۰ ۹۳.۵۲ ± ۰.۲۱ ۹۴.۱۳ ± ۰.۳۵ ۸۹.۳۲ ± ۰.۳۳

گوگل‌لینک-aux [ [30](#_bookmark45) ] 97.90 ± ۰.۳۴ ۹۳.۲۵ ± ۰.۳۳ ۹۳.۱۱ ± ۰.۵۵ ۸۹.۲۲ ± ۰.۲۵

EfficientNetB0-aux [ [30](#_bookmark45) ] 98.01 ± ۰.۴۵ ۹۳.۶۹ ± ۰.۱۱ ۹۳.۹۷ ± ۰.۱۳ ۸۹.۹۶ ± ۰.۲۷

EfficientNetB3-aux [ [30](#_bookmark45) ] 98.22 ± ۰.۴۹ ۹۴.۱۹ ± ۰.۱۵ ۹۴.۵۱ ± ۰.۷۵ ۹۱.۰۸ ± ۰.۱۴

|  |  |  |  |
| --- | --- | --- | --- |
| Proposed V32\_21k [384 × 384] 97.74 ± 0.10  Proposed V16\_21k [384 × 384] 98.49 ± 0.43 | 95.51 ± 0.57  95.86 ± 0.28 | 94.62 ± 0.38  95.56 ± 0.18 | 92.81 ± 0.17  92.60 ± 0.10  93.83 ± 0.46 |
| Proposed V16\_21k [384 × 384] 97.90 ± 0.10 | 94.27 ± 1.41 | 95.30 ± 0.58 | 93.05 ± 0.46 |

پیشنهادی V16\_21k [224 × ۲۲۴] ۹۸.۱۴ ± ۰.۴۷ ۹۴.۹۷ ± ۰.۰۱ ۹۵.۰۷ ± ۰.۱۲

[هرس کردن] ۵۰٪]

# نتیجه‌گیری

در این کار، ما روشی را برای طبقه‌بندی تصاویر سنجش از دور بر اساس ... ارائه کرده‌ایم . چشم انداز ترانسفورماتورها. متفاوت از سی‌ان‌ان‌ها، ‎‏ ... مدل است قادر به گرفتن دوربرد وابستگی‌ها در میان وصله‌ها از طریق یک توجه ماژول پیشنهادی روش بود روی چهار مجموعه داده تصویر سنجش از دور عمومی ارزیابی شد و نتایج تجربی نشان داد که اثربخشی از اینها جدید نوع از شبکه‌ها در در حال بهبود ‎‏ ... طبقه بندی دقت بالا در مقایسه با روش‌های پیشرفته علاوه بر این، ما نشان دادیم که با استفاده از ترکیبی از داده‌ها تقویت تکنیک‌ها می‌تواند کمک در بیشتر تقویت ‎‏ ... طبقه بندی دقت. به کاهش دادن ‎‏ ... اندازه از ‎‏ ... مدل، ما ارائه شده الف ساده فشرده‌سازی مدل راه حل که آلوها ‎‏ ... شبکه لایه‌ها برای آینده تحولات، ما پیشنهاد دادن تحقیق و تفحص رویکردهای جایگزین برای فشرده‌سازی ترانسفورماتور و تولید مدل‌های سبک‌وزن.

# منابع

**مشارکت‌های نویسنده:** YB این روش را طراحی و پیاده‌سازی کرد و مقاله را نوشت. LB، MMAR، RAD و NAA در تجزیه و تحلیل نتایج آزمایش و نگارش مقاله مشارکت داشتند. همه نویسندگان نسخه منتشر شده مقاله را خوانده و با آن موافقت کرده‌اند.

**تأمین مالی:** نویسندگان مراتب قدردانی خود را از طرح حمایت از پژوهشگران شماره (RSP-2020/69) دانشگاه ملک سعود، ریاض، عربستان سعودی ابراز می‌دارند.

**سپاسگزاری:** نویسندگان مراتب قدردانی خود را از پژوهشگران حامی پروژه ابراز می‌دارند. شماره (RSP-2020/69)، پادشاه سعود دانشگاه، ریاض، سعودی عربستان سعودی.

**درگیری‌ها از علاقه:**  نویسندگان اعلام کردن خیر درگیری از علاقه.

1. هو، کیو.؛ وو، دبلیو.؛ شیا، تی.؛ یو، کیو.؛ یانگ، پی.؛ لی، زی.؛ سونگ، کیو. بررسی استفاده از تصاویر گوگل ارث و روش‌های شیءگرا در نقشه‌برداری زمینی استفاده/پوشش نقشه برداری. *از راه دور سناتورهای* **۲۰۱۳** ، *۵* ، ۶۰۲۶–۶۰۴۲. [ [CrossRef](http://doi.org/10.3390/rs5116026) ]
2. توت، ج.؛ جو ó زاک ó دبلیو، جی. از راه دور حسگری پلتفرم‌ها و حسگرها: الف نظرسنجی *آی اس پی آر اس جی. فتوگرام. از راه دور سنس* **۲۰۱۶** ، *۱۱۵* ، ۲۲–۳۶.

[ [کراس رِف](http://doi.org/10.1016/j.isprsjprs.2015.10.004) ]

1. هوگندورن، اس پی؛ ون زویلن، اچ جی؛ شرودر، م.؛ گورت، ب.؛ وُسِلمن، جی. میکروسکوپی ترافیک داده‌ها مجموعه توسط از راه دور سنجش

*حمل و نقل پژوهش ضبط* **۲۰۰۳** ، *۱۸۵۵* ، ۱۲۱–۱۲۸. [ [CrossRef](http://doi.org/10.3141/1855-15) ]

1. والاوانیس، کی پی *ادونسز در بدون سرنشین هوایی وسایل نقلیه: ایالتی از ‎‏ ... هنر و ‎‏ ... جاده به آتونومی* ؛ رسانه علمی و تجاری اشپرینگر: برلین، آلمان، ۲۰۰۸؛ شابک ‎۹۷۸-۱-۴۰۲۰-۶۱۱۴-۱.
2. شپرد، ج.؛ رهنمون فر، م. زمان واقعی صحنه درک برای پهپاد تصویرسازی مبتنی بر روی عمیق کانولوشن عصبی شبکه‌ها. در مجموعه مقالات سمپوزیوم بین‌المللی علوم زمین و سنجش از دور IEEE 2017 (IGARSS)، فورت ورث، تگزاس، ایالات متحده آمریکا، 23-28 ژوئیه ۲۰۱۷؛ صفحات ۲۲۴۳–۲۲۴۶.
3. النجار، ح. کلانتر، بی. پرادان، بی. سعیدی، و. هالین، AA; اوئدا، ن. منصور، S. طبقه بندی پوشش زمین از ذوب شده تصاویر DSM و UAV با استفاده از شبکه‌های عصبی کانولوشن. *از راه دور سناتورها.* **۲۰۱۹** ، *۱۱* ، ۱۴۶۱. [ [CrossRef](http://doi.org/10.3390/rs11121461) ]
4. لیو، ت.؛ عبدالرحمن، ا.؛ زارع، ا.؛ دیویت، ب.آ.؛ فلوری، ل.؛ اسمیت، اس.ای. یک مدل شیءگرای زمینه‌محور کاملاً قابل یادگیری برای نقشه‌برداری پوشش زمین با استفاده از داده‌های چندنمایی از سیستم‌های هواپیمای بدون سرنشین. *از راه دور حس محیط.* **2018** ، *216* ، 328-344. [ [کراس رِف](http://doi.org/10.1016/j.rse.2018.06.031) ]
5. بازی، ی. شبکه عصبی دو شاخه‌ای برای یادگیری طبقه‌بندی چند برچسبی در تصاویر پهپاد. در مجموعه مقالات IGARSS 2019— ۲۰۱۹ IEEE بین‌المللی علوم زمین و از راه دور سنجش سمپوزیوم، یوکوهاما، ژاپن، ۲۸ ژوئیه–۲ مرداد ۲۰۱۹؛ ص. ۲۴۴۳–۲۴۴۶.
6. اسکیدمور، آ.ک. بیجکر، دبلیو. اشمیت، ک.؛ کومار، ل. استفاده کنید از از راه دور حسگری و سیستم اطلاعات جغرافیایی برای پایدار زمین مدیریت. *آی تی سی جی.* **۱۹۹۷** ، *۳* ،

۳۰۲–۳۱۵.

1. شیائو، وای.؛ ژان، کیو. مروری بر کاربردهای سنجش از دور در برنامه‌ریزی و مدیریت شهری در چین. در مجموعه مقالات رویداد مشترک سنجش از دور شهری ۲۰۰۹، شانگهای، چین، ۲۰ تا ۲۲ مه ۲۰۰۹؛ صفحات ۱ تا ۵.
2. دلدگان، جی ای؛ رابرتز، دی ای؛ د ریبیرو، اف اف طیفی مخلوط تحلیل در گوگل زمین موتور به مدل و ترسیم کردن آتش زخم‌هایی در گستره‌ای وسیع و در یک سری زمانی طولانی در منطقه انتقالی جنگل‌های بارانی-ساوانا. *Remote Sens. Environ.* **2019** , *232* , 111340. [ [کراس رِف](http://doi.org/10.1016/j.rse.2019.111340) ]
3. آهونن، ت.؛ حدید، الف.؛ پیتیکاینن، م. صورت توضیحات با محلی دودویی الگوها: کاربرد به صورت به رسمیت شناختن. *IEEE الگوی ترانس* *مقعدی ماخ هوش.* **۲۰۰۶** ، *۲۸* ، ۲۰۳۷–۲۰۴۱. [ [کراس رِف](http://doi.org/10.1109/TPAMI.2006.244) ]
4. دلال، ن.؛ تریگز، ب. هیستوگرام‌ها از گرا گرادیان ها برای انسان تشخیص. در مجموعه مقالات از ‎‏ ... ۲۰۰۵ IEEE کامپیوتر کنفرانس انجمن روی کامپیوتر چشم انداز و الگو شناخت (CVPR'05)، سان دیگو، کالیفرنیا، آمریکا، ۲۰–۲۵ ژوئن ۲۰۰۵؛ IEEE: سان دیگو، CA، ایالات متحده، 2005; جلد 1، ص 886–893.
5. لی، کیو.؛ کی، اس.؛ شن، وای.؛ نی، دی.؛ ژانگ، اچ.؛ وانگ، تی. هم‌ترازی تصویر چندطیفی با نقطه کلیدی غیرخطی مقیاس‌ناپذیر و بهبود یافته محلی ویژگی ماتریس. *IEEE علوم زمین. از راه دور سنس لت.* **۲۰۱۵** ، *۱۲* ، ۱۵۵۱–۱۵۵۵. [ [کراس رِف](http://doi.org/10.1109/LGRS.2015.2412955) ]
6. سیویک، جی.؛ راسل، پیش از میلاد؛ افروس، الف.آ؛ زیسرمن، الف.؛ فریمن، وزن خالص کشف کردن اشیاء و آنها مکان در تصاویر. در مجموعه مقالات ‎‏ ... دهم IEEE بین‌المللی کنفرانس روی کامپیوتر چشم انداز (ICCV'05)، پکن، چین، ۱۷–۲۱ اکتبر ۲۰۰۵؛ حجم ۱،

ص. ۳۷۰–۳۷۷.

1. هوانگ، ل.؛ چن، ج.؛ لی، دبلیو. دو، س. از راه دور حسگری تصویر صحنه طبقه بندی با استفاده از چند مقیاسی تکمیل شده محلی دودویی الگوها و بردارهای فیشر. *از راه دور سناتورها.* **2016** ، *8* ، 483. [ [CrossRef](http://doi.org/10.3390/rs8060483) ]
2. ایمبریاکو، ر.؛ سباستین، ج.؛ بوندارف، ای.؛ د با، پی اچ ان تجمیع شده عمیق محلی ویژگی‌ها برای از راه دور حسگری تصویر بازیابی.

*از راه دور سنس* **۲۰۱۹** ، *۱۱* ، ۴۹۳. [ [کراس رِف](http://doi.org/10.3390/rs11050493) ]

1. دیائو، دبلیو.؛ سان، ایکس.؛ ژنگ، ایکس.؛ دو، اف.؛ وانگ، اچ.؛ فو، کی. تشخیص کارآمد اشیاء مبتنی بر برجستگی در تصاویر سنجش از دور با استفاده از داده‌های عمیق باور شبکه‌ها. *IEEE علوم زمین. از راه دور سنس لت.* **۲۰۱۶** ، *۱۳* ، ۱۳۷–۱۴۱. [ [CrossRef](http://doi.org/10.1109/LGRS.2015.2498644) ]
2. چن، ی.؛ لین، ز.؛ ژائو، ایکس. وانگ، جی.؛ گو، ی. عمیق مبتنی بر یادگیری طبقه بندی از فراطیفی داده‌ها. *IEEE جی. سل بالا. کاربرد مشاهدات زمین* *. از راه دور سنس* **۲۰۱۴** ، *۷* ، ۲۰۹۴–۲۱۰۷. [ [CrossRef](http://doi.org/10.1109/JSTARS.2014.2329330) ]
3. نوگیرا، ک. میراندا، وو؛ سانتوس، جاد بهبود فضایی ویژگی نمایندگی از هوایی صحنه‌ها توسط با استفاده از شبکه‌های کانولوشنی در مجموعه مقالات بیست و هشتمین کنفرانس SIBGRAPI 2015 در مورد گرافیک، الگوها و تصاویر، سالوادور، برزیل، 26 تا 29 آگوست 2015؛ صفحات 289 تا 296.
4. مارمانی‌ها، دی. داتچو، م.؛ اش، تی.؛ استیلا، یو. عمیق یادگیری زمین مشاهده طبقه بندی با استفاده از imagenet از پیش آموزش دیده شبکه‌ها.

*IEEE علوم زمین. از راه دور سناتور لت.* **۲۰۱۶** ، *۱۳* ، ۱۰۵–۱۰۹. [ [کراس رِف](http://doi.org/10.1109/LGRS.2015.2499239) ]

1. ماگیوری، ای. تارابالکا، ی. چارپیات، جی. Alliez, P. شبکه های عصبی کانولوشن برای طبقه بندی تصاویر سنجش از دور در مقیاس بزرگ بخش *IEEE ترانس. علوم زمین. از راه دور سنس* **۲۰۱۷** ، *۵۵* ، ۶۴۵–۶۵۷. [ [کراس رِف](http://doi.org/10.1109/TGRS.2016.2612821) ]
2. لاخال، میشیگان؛ چویکالپ، ه.؛ اسکالرا، س.؛ اوفلی، ف. مکرر شبکه‌های عصبی برای کنترل از راه دور تصویر حسگر طبقه بندی. *IET Comput.* *Vis.* **2018** ، *12* ، 1040–1045. [ [CrossRef](http://doi.org/10.1049/iet-cvi.2017.0420) ]
3. ژو، ل.؛ چن، ی.؛ قمیسی، پ.؛ بندیکتسون، ج.ا. شبکه‌های مولد تخاصمی برای طبقه‌بندی تصاویر فراطیفی. *IEEE* *Trans. علوم زمین. از راه دور سنس* **۲۰۱۸** ، *۵۶* ، ۵۰۴۶–۵۰۶۳. [ [کراس رِف](http://doi.org/10.1109/TGRS.2018.2805286) ]
4. فنگ، جی.؛ یو، ح.؛ وانگ، ل.؛ کائو، ایکس. ژانگ، ایکس. جیائو، ل. طبقه‌بندی از فراطیفی تصاویر مبتنی بر روی چندکلاسی فضایی-طیفی شبکه‌های مولد تخاصمی. *IEEE Trans. Geosci. Remote Sens.* **2019** , *57* , 5329–5343. [ [CrossRef](http://doi.org/10.1109/TGRS.2019.2899057) ]
5. مو، ل.؛ لو، ایکس. لی، ایکس. ژو، بیست و یک غیر محلی گراف کانولوشن شبکه‌ها برای فراطیفی تصویر طبقه بندی. *IEEE Trans.* *Geosci. Remote سناتورها.* **۲۰۲۰** ، ۱–۱۲. [ [CrossRef](http://doi.org/10.1109/TGRS.2020.2973363) ]
6. هو، دبلیو. لی، ح.؛ پان، ل.؛ لی، دبلیو. تائو، ر.؛ دو، س. فضایی-طیفی ویژگی استخراج از طریق عمیق ConvLSTM عصبی شبکه‌ها برای طبقه‌بندی تصاویر فراطیفی. *IEEE ترانس. علوم زمین. از راه دور سناتورها.* **2020** ، *58* ، 4237-4250. [ [CrossRef](http://doi.org/10.1109/TGRS.2019.2961947) ]
7. بی، س.؛ شین، ک.؛ لی، ز.؛ ژانگ، ح.؛ خو، ک.؛ شیا، جی.-اس. الف چند نمونه‌ای دارای اتصالات متراکم شبکه کانولوشن برای هوایی صحنه طبقه بندی.

*IEEE ترانس. تصویر فرآیند.* **۲۰۲۰** ، *۲۹* ، ۴۹۱۱–۴۹۲۶. [ [کراس رِف](http://doi.org/10.1109/TIP.2020.2975718) ]

1. یو، ی.؛ لی، ایکس. لیو، ف. توجه به شبکه‌های مولد تخاصمی (GAN): یادگیری عمیق بدون نظارت برای طبقه‌بندی صحنه هوایی *انتقال IEEE. علوم زمین.* *از راه دور سناتورها.* **۲۰۲۰** ، *۵۸* ، ۵۱۹–۵۳۱. [ [CrossRef](http://doi.org/10.1109/TGRS.2019.2937830) ]
2. بازی، ی.؛ آل رحال، میلی متر؛ الهیچری، ح.؛ آلاجلان، ن. ساده هنوز مؤثر تنظیم دقیق از عمیق سی‌ان‌ان‌ها با استفاده از یک کمکی طبقه بندی از دست دادن برای طبقه‌بندی صحنه‌های سنجش از دور. *سنجش از دور.* **2019** ، *11* ، 2908. [ [CrossRef](http://doi.org/10.3390/rs11242908) ]
3. خورشید، ح.؛ لی، س.؛ ژنگ، ایکس. لو، ایکس. از راه دور حسگری صحنه طبقه بندی توسط دروازه دار دو جهته شبکه. *IEEE علوم زمین.* *سنجش از دور.* **2019** ، 1-15. [ [CrossRef](http://doi.org/10.1109/TGRS.2019.2931801) ]
4. لیو، ی.؛ لیو، ی.؛ دینگ، ل. طبقه‌بندی صحنه بر اساس ادغام عمیق ویژگی‌های دو مرحله‌ای. *IEEE Geosci. حسگرهای از راه دور لت.* **2018** ، *15* ، 183–186. [ [CrossRef](http://doi.org/10.1109/LGRS.2017.2779469) ]
5. یو، ی.؛ لیو، ف. الف دو جریانی عمیق فیوژن چارچوب برای وضوح بالا هوایی صحنه طبقه بندی. موجود است آنلاین: <https://www.hindawi.com/journals/cin/2018/8639367/> (دسترسی پیدا کرد روی ۲۰ نوامبر ۲۰۲۰).
6. چنگ، جی.؛ یانگ، سی. یائو، X. گوا، ل.؛ هان، جی. چه زمانی عمیق یادگیری ملاقات می‌کند متریک یادگیری: از راه دور حسگری تصویر صحنه طبقه بندی از طریق یادگیری تبعیض آمیز CNNها. *IEEE ترانس. علوم زمین. از راه دور سنس* **۲۰۱۸** ، *۵۶* ، ۲۸۱۱–۲۸۲۱. [ [CrossRef](http://doi.org/10.1109/TGRS.2017.2783902) ]
7. شو، دبلیو. دای، ایکس. لیو، ل. از راه دور حسگری صحنه طبقه بندی مبتنی بر روی چند سازه عمیق ویژگی‌ها همجوشی *IEEE دسترسی* **۲۰۲۰** ، *۸* ،

۲۸۷۴۶–۲۸۷۵۵. [ [کراس رِف](http://doi.org/10.1109/ACCESS.2020.2968771) ]

1. وانگ، کیو.؛ لی، بی.؛ شیائو، تی.؛ ژو، جی.؛ لی، سی.؛ وانگ، دی. اف.؛ چائو، ال. اس. یادگیری مدل‌های مبدل عمیق برای ترجمه ماشینی. در مجموعه مقالات از ‎‏ ... پنجاه و هفتم سالانه جلسه از ‎‏ ... انجمن برای محاسباتی زبان‌شناسی، فورتزا دا فلورانس، ایتالیا، ۲۸ ژوئیه–۲ اوت ۲۰۱۹؛ صفحات ۱۸۱۰–۱۸۲۲.
2. دای، ز.؛ یانگ، ز.؛ یانگ، ی.؛ کاربونل، ج.؛ لی، کیو.؛ سالاخوتدینوف، ر. Transformer-XL: مدل‌های زبانیِ توجه‌گرا فراتر از یک بافت با طول ثابت. در مجموعه مقالات پنجاه و هفتمین نشست سالانه انجمن زبان‌شناسی محاسباتی، فورتزا دا فلورانس، ایتالیا، ۲۸ ژوئیه–۲ مرداد ۲۰۱۹؛ صفحات ۲۹۷۸–۲۹۸۸.
3. چن، ن.؛ واتانابه، س.؛ ویلالبا، ج.ا.؛ زلاسکو، پ.؛ دهاک، ن. مبدل غیر خودهمبسته برای تشخیص گفتار. *فرآیند سیگنال IEEE . لت.* **2020** ، *28* ، 121–125. [ [CrossRef](http://doi.org/10.1109/LSP.2020.3044547) ]
4. واسوانی، الف.؛ شازیر، ن.؛ پارمار، ن.؛ اوزکوریت، جی.؛ جونز، ل.؛ گومز، یک؛ قیصر، ل.؛ پولوسوخین، من. توجه است همه شما نیاز. *پیشرفته.* *عصبی اطلاعات فرآیند. سیستم* **۲۰۱۷** ، *30* ، ۵۹۹۸–۶۰۰۸.
5. بلو، آی.؛ زوف، بی.؛ واسوانی، ای.؛ شلنز، جی.؛ لی، کیو.وی. شبکه‌های کانولوشنی تقویت‌شده با توجه. در مجموعه مقالات IEEE/CVF 2019 بین‌المللی کنفرانس روی کامپیوتر چشم انداز (ICCV) سئول، کره، ۲۷ اکتبر–۲ نوامبر ۲۰۱۹؛ صفحات ۳۲۸۵–۳۲۹۴.
6. وانگ، کیو.؛ لیو، اس.؛ چانوسوت، جی.؛ لی، ایکس. طبقه‌بندی صحنه با توجه مکرر تصاویر سنجش از دور VHR. *IEEE علوم زمین شناسی ترانس .* *از راه دور سنس* **۲۰۱۹** ، *۵۷* ، ۱۱۵۵–۱۱۶۷. [ [CrossRef](http://doi.org/10.1109/TGRS.2018.2864987) ]
7. وو، بی. خو، سی. دای، ایکس. وان، آ. ژانگ، پی. تومیزوکا، م. کوتزر، ک. Vajda، P. ترانسفورماتورهای بصری: نمایش تصویر مبتنی بر رمز و پردازش برای کامپیوتر بینایی. *آرشیو* **۲۰۲۰** ، arXiv:2006.03677.
8. راماچاندران، پ.؛ پارمار، ن.؛ واسوانی، الف.؛ بلو، من.؛ لِوِسکایا، الف.؛ شلنز، جی. خود-توجهی مستقل در بینایی مدل‌ها. *آرشیو*

**۲۰۱۹** ، arXiv:1906.05909.

1. چن، م.؛ رادفورد، الف.؛ کودک، ر.؛ وو، جی.؛ جون، ح.؛ لوان، دی. ساتسکور، من. مولد پیش آموزش از پیکسل‌ها در مجموعه مقالات از ‎‏ ... سی و هفتمین کنفرانس بین‌المللی یادگیری ماشین، وین، اتریش، ۱۲ تا ۱۸ ژوئیه ۲۰۲۰؛ صفحات. ۱۶۹۱–۱۷۰۳.
2. دوسوویتسکی، آ. بیر، ال. کولسنیکوف، آ. وایسنبورن، دی. ژای، ایکس. Unterthiner، T. دهقانی، م. مایندرر، م. هیگلد، جی. ژلی، س.؛ و ال. آن تصویر است ارزش ۱۶ ۱۶ کلمات کلیدی: ترانسفورماتورها برای تصویر به رسمیت شناختن در مقیاس. *arXiv* **۲۰۲۰** ، arXiv:2010.11929.

×

1. هِی، جی.؛ ژائو، ال.؛ یانگ، اچ.؛ ژانگ، ام.؛ لی، دبلیو. HSI-BERT: طبقه‌بندی تصاویر فراطیفی با استفاده از نمایش رمزگذار دوجهته از ترانسفورماتورها. *IEEE ترانس. علوم زمین. از راه دور سناتورها.* **۲۰۲۰** ، *۵۸* ، ۱۶۵–۱۷۸. [ [CrossRef](http://doi.org/10.1109/TGRS.2019.2934760) ]
2. دولین، جی.؛ چانگ، ام.-دبلیو.؛ لی، کی.؛ توتانووا، کی. برت: پیش‌آموزش مبدل‌های دوجهته عمیق برای درک زبان. در مجموعه مقالات کنفرانس سال ۲۰۱۹ شعبه آمریکای شمالی انجمن زبان‌شناسی محاسباتی: زبان انسانی زبان فناوری‌ها، مینیاپولیس، ام آی، آمریکا، ۲–۷ ژوئن ۲۰۱۹؛ طولانی و کوتاه کاغذها حجم ۱، ص. ۴۱۷۱–۴۱۸۶.
3. کوبک، ای‌دی؛ زوف، بی.؛ مانه، دی.؛ واسودوان، وی.؛ لی، کیو. وی. خودافزایی: راهبردهای تقویت یادگیری از داده‌ها. در مجموعه مقالات کنفرانس IEEE/CVF 2019 در مورد بینایی کامپیوتر و تشخیص الگو (CVPR)، لانگ بیچ، کالیفرنیا، ایالات متحده آمریکا، 15 تا 21 ژوئن 2019؛ IEEE: لانگ بیچ، کالیفرنیا، ایالات متحده آمریکا، 2019؛ صفحات 113 تا 123.
4. جکسون، پی.تی؛ عطاپور-ابرغویی، ای.؛ بونر، اس.؛ برکون، تی.پی؛ اوبارا، بی. تقویت سبک: تقویت داده‌ها از طریق تصادفی‌سازی سبک. در مجموعه مقالات از ‎‏ ... ۲۰۱۹ IEEE کنفرانس روی کامپیوتر چشم انداز و الگو شناخت (CVPR) طولانی بیچ، کالیفرنیا، ایالات متحده آمریکا، ۱۶ تا ۲۰ ژوئن ۲۰۱۹.
5. بولز، سی.؛ چن، ال.؛ گررو، آر.؛ بنتلی، پی.؛ گان، آر.؛ هامز، ای.؛ دیکی، دی.ای.؛ هرن آندز ، ام.وی.؛ واردلا، جی.؛ روکرت، دی. تقویت GAN: تقویت داده‌های آموزشی با استفاده از شبکه‌های مولد تخاصمی. *arXiv* **۲۰۱۸** ، arXiv:1810.10863.
6. دی‌وریس، تی.؛ تیلور، جی دبلیو بهبود یافته منظم سازی از کانولوشن عصبی شبکه‌ها با بریدگی *آرشیو* **۲۰۱۷** ، arXiv:1708.04552.
7. ژانگ، ح.؛ سیسه، م.؛ دوفین، ی.ن؛ لوپز-پاز، دی. اختلاط: فراتر از تجربی ریسک کوچک‌سازی *آرکایو* **۲۰۱۸** ، arXiv:1710.09412.
8. یون، س.؛ هان، د.؛ چون، س.؛ اوه، س.ج.؛ یو، ی.؛ چو، ج. CutMix: استراتژی منظم‌سازی برای آموزش طبقه‌بندی‌کننده‌های قوی با ویژگی‌های قابل محلی‌سازی. در مجموعه مقالات کنفرانس بین‌المللی IEEE/CVF 2019 در مورد بینایی کامپیوتر (ICCV)، سئول، کره، 27 اکتبر تا 2 نوامبر ۲۰۱۹؛ صفحات ۶۰۲۲–۶۰۳۱.
9. هینتون، جی.؛ وینیالز، اُ.؛ دین، جی. تقطیر ‎‏ ... دانش در الف عصبی شبکه. *آرشیو* **۲۰۱۵** ، arXiv:1503.02531.
10. هان، س.؛ مائو، ه.؛ دالی، دبلیو.جی. فشرده‌سازی عمیق: فشرده‌سازی شبکه‌های عصبی عمیق با هرس کردن، کوانتیزاسیون آموزش‌دیده و کدگذاری هافمن. *arXiv* **2016** ، arXiv:1510.00149.
11. وو، جی.؛ لنگ، ج.؛ وانگ، ی.؛ هو، س.؛ چنگ، جی. کوانتیزه شده کانولوشن عصبی شبکه‌ها برای موبایل دستگاه‌ها. در مجموعه مقالات از کنفرانس IEEE 2016 در مورد بینایی کامپیوتر و تشخیص الگو (CVPR)، لاس وگاس، نوادا، ایالات متحده آمریکا، 27 تا 30 ژوئن 2016؛ صفحات. ۴۸۲۰–۴۸۲۸.
12. یانگ، ی.؛ نیوسام، س. مجموعه کلمات بصری و بسط‌های مکانی برای طبقه‌بندی کاربری زمین. در مجموعه مقالات هجدهمین SIGSPA- تیال بین‌المللی کنفرانس روی پیشرفت‌ها در جغرافیایی اطلاعات سیستم‌ها—GIS سال دهم، سان خوزه، کالیفرنیا، آمریکا، ۲–۵ نوامبر ۲۰۱۰؛ ص ۲۷۰.
13. شیا، جی.؛ هو، جی.؛ هو، ف. شی، ب.؛ بای، ایکس. ژونگ، ی.؛ ژانگ، ل.؛ لو، ایکس. کمک: الف معیار داده‌ها مجموعه برای عملکرد ارزیابی از هوایی طبقه‌بندی صحنه. *IEEE Trans. Geosci. Remote Sens.* **2017** , *55* , 3965–3981. [ [CrossRef](http://doi.org/10.1109/TGRS.2017.2685945) ]
14. او، ن.؛ نیش، ل.؛ لی، س.؛ پلازا، الف.؛ پلازا، جی. از راه دور حسگری صحنه طبقه بندی با استفاده از چند لایه انباشته شده کوواریانس جمع کردن *IEEE* *Trans. Geosci. Remote Sens.* **2018** , *56* , 6899–6910. [ [CrossRef](http://doi.org/10.1109/TGRS.2018.2845668) ]
15. چنگ، جی.؛ هان، جی.؛ لو، ایکس. از راه دور حسگری تصویر صحنه طبقه بندی: معیار و ایالت از ‎‏ ... هنر. *مجموعه مقالات IEEE* **2017** ، *۱۰۵* ، ۱۸۶۵–۱۸۸۳. [ [CrossRef](http://doi.org/10.1109/JPROC.2017.2675998) ]