Archive of SID.ir

سنجش از دور
و
GIS ایران

سنجش از دور و GIS ایران سال پانزدهم، شماره سوم، پاییز ۱۴۰۲
Iranian Remote Sensing & GIS Vol. ۱۵, No. ۳, Autumn ۲۰۲۳

۴۱-۶۰
مقاله پژوهشی

بهبود طبقه‌بندی تصاویر ابرطیفی با استفاده از مدل ترکیبی شبکه‌های کپسول و درخت تصمیم تقویتی

پویا احمدی[۱*]، طیبه مناقبی[۲]، حمید عبادی[۳]، بهنام اصغری بیرامی[۴]

۱. دانشجوی دکتری سنجش از دور، گروه مهندسی نقشه‌برداری و اطلاعات مکانی، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران
۲. استادیار دانشکده مهندسی نقشه برداری و اطلاعات مکانی، دانشگاه تهران، تهران، ایران
۳. استاد گروه مهندسی فتوگرامتری و سنجش ازدور، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران
۴. دکتری سنجش از دور، گروه مهندسی فتوگرامتری و سنجش از دور، دانشکده مهندسی نقشه برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران

تاریخ دریافت مقاله: ۱۴۰۰/۰۸/۱۰ تاریخ پذیرش مقاله: ۱۴۰۱/۰۱/۲۹

چکیده

با گسترش دانش سنجش از دور، استفاده از تصاویر هایپراسپکترال روزبه‌روز افزایش و عمومیت می‌یابد. طبقه‌بندی یکی از محبوب‌ترین موضوعات در سنجش از دور ابرطیفی است. طی دو دهۀ گذشته، روش‌های بسیاری برای مقابله با مشکل طبقه‌بندی داده‌های هایپراسپکترال پیشنهاد شده است. در پژوهش حاضر، ساختاری مبتنی‌بر یادگیری شبکه‌های کپسول برای طبقه‌بندی تصاویر ابرطیفی به‌کار رفته است؛ به‌گونه‌ای که ساختار شبکه بتواند، با استفاده از یک لایۀ کانولوشنی و یک لایۀ کپسول، بهترین حالت تولید ویژگی‌ها را داشته باشد و درعین‌حال از بیش‌برازش شبکه روی نمونه‌های آموزشی جلوگیری کند. نتایج به‌دست‌آمده نشان از کیفیت بالای ویژگی‌های تولیدی در ساختار پیشنهادی دارد. درراستای بهبود دقت طبقه‌بندی، رویکرد استخراج ویژگی ازطریق شبکۀ طراحی‌شده و طبقه‌بندی با استفاده از الگوریتم درخت تقویتی XGBoost، با روش طبقه‌بندی ازطریق شبکۀ عمیق سراسری مقایسه شد تا، علاوه‌بر بررسی و کیفیت‌سنجی ویژگی‌های عمیق برداری تولیدی به‌روش پیشنهادی در طبقه‌بندی‌کننده‌های گوناگون، میزان توانایی شبکه‌های عمیق سراسری نیز، در کاربرد طبقه‌بندی، بررسی شود. رویکرد کپسول پیشنهادی شامل سه لایۀ اصلی است: ۱) Prime با کپسول‌هایی به‌اندازۀ ۸ و ۳۲ فیلتر ۹×۹ و گام حرکتی ۲؛ ۲) Digitcaps دارای دَه کپسول شانزده‌بعدی؛ ۳) لایۀ تماماً متصل. نتایج بررسی دو رویکرد برای شبکۀ عمیق و نیز ترکیب شبکه‌های کپسول با الگوریتم درخت تقویتی XGBoost مقایسه شد. رویکردهایی همچون SVM، RF-۲۰۰، LSTM، GRU، و GRU-Pretanh برای مقایسۀ رویکرد پیشنهادی براساس پیکربندی‌هایی درنظر گرفته شدند که در تحقیقات به آنها اشاره شده بود. برای ارزیابی مدل پیشنهادی، مجموعه دادۀ Indian Pines نیز، شامل شانزده کلاس متفاوت، به‌کار رفت. با استفاده از روش پیشنهادی ترکیبی، طبقه‌بندی تصاویر با دقت ۹۹٪ روی‌داده‌های آموزش و دقت ۹۷/۵٪ روی داده‌های تست انجام می‌شود.

کلید واژه‌ها: طبقه‌بندی، تصاویر ابرطیفی، شبکه‌های کپسول، درخت تصمیم تقویتی، مدل ترکیبی.

* نویسندۀ مکاتبه‌کننده: تهران، خیابان ولیعصر، دانشگاه صنعتی خواجه نصیرالدین طوسی، دانشکدۀ مهندسی نقشه‌برداری و اطلاعات مکانی. تلفن: ۰۹۱۸۵۳۶۳۶۰۴
Email: ahmadi.pouya@ut.ac.ir
https://orcid.org/۰۰۰۹-۰۰۰۱-۷۳۰۸-۱۰۵٤
https://dx.doi.org/۱۰,٤۸۳۰۸/gisj.۲۰۲۳,۱۰۲۳٤۷

Copyright: © ۲۰۲۳ by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/۴.۰/).

Archive of SID.ir

Archive of SID.ir

۱- مقدمه

سنجنده‌های ابرطیفی[1] نوع خاصی از سنجنده‌های طیف‌سنجی‌اند که پهنای باند مورد نظر را به صدها باند باریک مجاور هم تقسیم می‌کنند. دردسترس قرارگرفتن سنجنده‌هایی، با توان تفکیک طیفی[2] بالا و تعداد باند زیاد، دستیابی به جزئیات بیشتر درمورد کلاس‌ها را امکان‌پذیر کرده و باعث افزایش دقت و اعتبار طبقه‌بندی شده است. با توجه به تعداد بسیار باندها، تصاویر ابرطیفی کاربردهای فراوانی می‌یابند (Ghassemian et al., 2003). استفاده از تصاویر ابرطیفی در کاربردهایی مانند کشاورزی و برآورد دقیق محصولات، جنگل‌داری و مطالعهٔ پوشش‌های گیاهی، مطالعهٔ منابع آب و شناسایی آلودگی‌ها، اکتشاف منابع طبیعی و معدنی، زمین‌شناسی و نفت، کاربردهای صلح‌آمیز همچون شناسایی و پاک‌سازی مناطق جنگی آلوده، کاربردهای نظامی و شناسایی ادوات استتارشده نتایج مطلوبی به‌دست آورده است.

مهم‌ترین چالش، در پردازش تصاویر ابرطیفی، اندک‌بودن تعداد نقاط آموزشی به‌نسبت تعداد باندهای تصویر است. درواقع، مسئله این است که اگر تعداد نقاط آموزشی به‌نسبت تعداد باندهای داده، کمتر از یکصد نمونه به‌ازای هر کلاس باشد، با افزایش تعداد باندها، خطای تخمین پارامترهای طبقه‌بندی‌کننده بیشتر می‌شود (Keshavarz & Ghasemiyan, 2005). با توجه به این مسئله، یک راه‌حل برای بهبود عملکرد طبقه‌بندی‌کننده‌های آماری، کاهش فضای ویژگی پیش‌اَز طبقه‌بندی این تصاویر است که به افزایش نسبت تعداد نقاط آموزشی به تعداد ویژگی منجر می‌شود. برای این منظور، دو دسته الگوریتم به‌کار می‌رود: الگوریتم‌های انتخاب باند[3] و الگوریتم‌های کاهش ابعاد فضای ویژگی[4]. در سال‌های اخیر، الگوریتم‌های متعددی برای استخراج ویژگی مطرح و استفاده شده‌اند. استخراج ویژگی با مرز تصمیم‌گیری[5] (Lee & Landgrebe, 1993)، نگاشت جست‌وجوگر (Jimenez & Landgrebe, 1999)، تحلیل ممیز و تبدیل موجک[6] (Kaewpijit et al., 2003)، تحلیل مؤلفه‌های اصلی[7]، تحلیل مؤلفه‌های مستقل[8] (Du et al., 2008) ازجمله الگوریتم‌هایی‌اند که به‌منظور کاهش ابعاد فضای ویژگی و بهبود طبقه‌بندی تصاویر ابرطیفی به‌کار رفته‌اند. بین سال‌های ۱۹۹۸ تا ۲۰۰۰، گوالتیری و چِتری[9] (۲۰۰۰) کارآیی ماشین‌های بردار پشتیبان[10] در طبقه‌بندی تصاویر ابرطیفی را بررسی کردند و نشان دادند این روش امکان آنالیز تصاویر ابرطیفی را در فضای ویژگی با ابعاد زیاد و بدون نیاز به فرایند کاهش ویژگی، فراهم می‌کند. ملگانی و بروزونه[11] (۲۰۰۴) نیز کارآیی ماشین‌های بردار پشتیبان در طبقه‌بندی تصاویر ابرطیفی را مطالعه کردند. لیانگ[12] و همکاران (۲۰۰۵) ماشین بردار رگرسیون[13] را برای طبقه‌بندی تصاویر ابرطیفی پیشنهاد دادند. از دید آنها، پیش‌بینی‌های احتمالی، تخمین اتوماتیک پارامترها و امکان انتخاب هسته‌های اختیاری از مزایای RVM بر SVM است. فوکوناگا و اسبر[14] (۲۰۰۸) مشکل اندک‌بودن تعداد نمونه‌های آموزشی در طبقه‌بندی تصاویر ابرطیفی را بررسی و نشان دادند تعداد نقاط آموزشی مورد نیاز برای طبقه‌بندی‌کننده‌های خطی، متناسب با افزایش تعداد باندها، به‌صورت خطی افزایش می‌یابد. بروزونه و همکاران (۲۰۰۶) روش TSVM[15] را برای طبقه‌بندی داده‌هایی مطرح کردند که نمونه‌های آموزشی آنها کاملاً مورد اعتماد نیستند. فاول[16] و همکاران (۲۰۰۷) استفادهٔ

1. Hyperspectral
2. Spectral Resolution
3. Band Selection
4. Dimension reduction
5. Feature Extraction Based on Decision Boundaries
6. Automatic Wavelet Spectral Analysis
7. Principal Component Analysis
8. Independent Component Analysis
9. Gualtieri & Chetri
10. Suport Vector Machin
11. Melgani & Bruzzone
12. Liyang
13. Regression Vector Machin
14. Fukunaga & Sber
15. Transductive SVM
16. Fauvel

Archive of SID.ir

Archive of SID.ir

هم‌زمان از ویژگی‌های طیفی و مکانی در طبقه‌بندی تصاویر ابرطیفی به‌روش SVM را بررسی کردند. نتایج این بررسی روی تصویر ROSIS مؤید افزایش کم‌دقت طبقه‌بندی، بدون کاهش ابعاد فضای ویژگی و بهبود چشمگیر دقت، با استفاده از الگوریتم کاهش ابعاد فضای ویژگی بود. اکبری[1] و همکاران (۲۰۱۶) با استخراج اطلاعات از پیکسل‌های همسایگی در فرایند انتخاب نمونه‌ها، کوشیدند دقت الگوریتم‌های طبقه‌بندی طیفی- مکانی در تصاویر ابرطیفی را بهبود ببخشند. غفاری[2] و همکاران (۲۰۱۶) انتخاب باندهای بهینه به‌منظور بهبود جداسازی طیفی تصاویر ابرطیفی را بررسی کردند. آنها روش جدیدی برای انتخاب باند متشکل از اولویت‌بندی‌ها، درراستای کاهش اثر تغییرپذیری طیفی و کاهش همبستگی بین باندها مطرح کردند و نشان دادند می‌توان به حذف مؤثر بسیاری از باندهای غیرضروری اقدام کرد.

در چند سال اخیر، شبکه‌های عصبی کانولوشن (CNN)[3] به‌منزلهٔ روش منتخب برمبنای یادگیری عمیق[4] برای تولید ویژگی از تصاویر، درنظر گرفته شده است زیرا ساختاری مناسب برای کار با تصاویر دارد (Ghamisi et al., 2017). نکتهٔ مهم دراین‌زمینه تعیین ساختار شبکهٔ متناسب با داده‌های ورودی به شبکه است. لوئو[5] (۲۰۱۸) تلاش داشت از شبکه‌های کپسول[6] استفاده کند؛ وی رویکردی مبتنی‌بر CNN را با مدل مبتنی‌بر CapsNet مقایسه کرد. نتایج این مطالعه مؤید آن بود که روش مبتنی‌بر CapsNet مزایای مورد انتظار را فراهم نمی‌کند و هیچ بینش عمیقی نشان داده نشده است. وانگ[7] و همکاران (۲۰۱۸) یک ساختار CapsNet براساس شبکه‌های عصبی بازگشتی (RNN)[8] را به‌منظور تجزیه‌وتحلیل احساسات و عواطف مطرح کردند. در پژوهشی دیگر، شبکه‌های عصبی کپسول این‌گونه تعریف شد که نوع کاملاً جدیدی از معماری یادگیری عمیق را نشان می‌دهد که قصد دارد بر محدودیت‌ها و اشکالات CNN، همچون ازدست‌رفتن اطلاعات در مرحلهٔ Max Pooling غلبه کند (Sabour et al., 2017).

هو[9] و همکاران (۲۰۱۵) شبکهٔ کانولوشنی عصبی پنج‌لایه‌ای برای داده‌های ابرطیفی ارائه دادند. در این روش، از اطلاعات طیفی بدون درنظرگرفتن اطلاعات مکانی استفاده شد. یو[10] و همکاران (۲۰۱۹) شبکهٔ کانولوشنی کپسول را برای تشخیص وسایل نقلیه، با استفاده از تصاویر سنجش از دور پیشنهاد دادند. در رویکرد پیشنهادی، تصویر سنجش از دور با استفاده از استراتژی قطعه‌بندی سوپرپیکسل[11] به چندین تکه[12] تقسیم شد و به‌صورت ورودی برای تشخیص وجود وسایل نقلیه به‌کار رفت. گائو[13] و همکاران (۲۰۲۰)، رویکرد شبکهٔ عصبی کپسول چندمقیاسه را برای تشخیص تغییرات از تصاویر رادار با دهانهٔ مصنوعی[14] طراحی کردند. در این مطالعه، مدل کپسول صرفاً برای مشخص‌کردن رابطهٔ بین ویژگی‌ها استفاده شد و یک انطباقی کانولوشن، برای افزایش قدرت شناسایی رابطهٔ بین ویژگی‌ها به‌کار رفت. در ادامه، درمورد شبکه‌های عصبی کانولوشن سه‌بعدی (CNN)، به‌منظور طبقه‌بندی تصاویر ابرطیفی با استفادهٔ هم‌زمان از ویژگی‌های طیفی و مکانی، بحث و بررسی شد و درنَهایت مدل شبکهٔ کپسول سه‌بعدی، با استفاده از کپسول‌های طیفی- مکانی، جایگزین شبکه‌های کانولوشنی شدند (Paoletti et al., 2018; 2019). شو[15] و همکاران (۲۰۲۰) یک شبکهٔ عصبی کپسول با شبکهٔ کانولوشنی اکتاو[16] را برای طبقه‌بندی

1. Akbari
2. Ghaffari
3. Convolution Neural Networks
4. Deep Learning
5. Luo, Zou
6. CapsNet
7. Wang
8. Recurrent Neural Network
9. Hu
10. Yu
11. Superpixel Segmentation Strategy
12. Patch
13. Gao
14. Synthetic Aperture Radar
15. Xu
16. Octave Convolution

Archive of SID.ir

Archive of SID.ir

تصاویر ابرطیفی، طراحی کردند. در رویکرد طراحی‌شده، شبکهٔ کانولوشنی معماری‌های موازی را سازگار و ویژگی‌های تصاویر را استخراج کرد.

در این پژوهش، از ساختاری مبتنی‌بر یادگیری عمیق برای طبقه‌بندی تصاویر ابرطیفی استفاده شده است؛ به‌گونه‌ای که بتوان از ویژگی‌های معماری شبکهٔ کپسول برای رفع معایب شبکهٔ کانولوشنی، به‌منظور طبقه‌بندی تصاویر ابرطیفی و مقایسهٔ نتایج این دو شبکه، استفاده کرد. در این مطالعه، با توجه به اینکه خروجی‌های شبکهٔ کپسول به‌صورت بردار است، می‌توان ارتباط بین ویژگی‌های استخراجی را ازطریق اندازه و جهت بردار، استخراج کرد؛ هریک از المان‌های بردار خروجی نشان‌دهندهٔ توصیف خاصی از موجودیت تصویر است. رویکرد استخراج ویژگی، ازطریق شبکهٔ طراحی‌شده و طبقه‌بندی با استفاده از الگوریتم درخت تقویتی XGBoost[1] با روش طبقه‌بندی ازطریق شبکهٔ عمیق سراسری مقایسه شده است تا، علاوه‌بر بررسی و کیفیت‌سنجی ویژگی‌های عمیق تولیدی به‌روش پیشنهادی در طبقه‌بندی‌کننده‌های متفاوت، میزان توانایی شبکه‌های عمیق سراسری نیز در کاربرد طبقه‌بندی، بررسی شود.

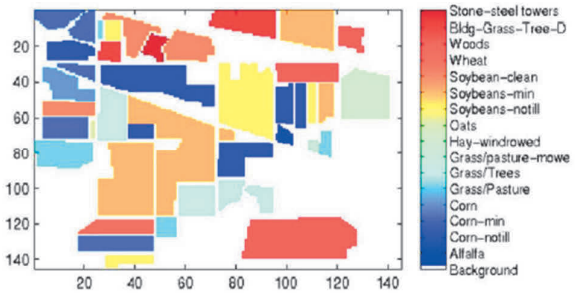
۲- داده‌ها و معرفی منطقهٔ مورد مطالعه

تصاویر ابرطیفی در جامعهٔ سنجش از دور، به‌منظور استفاده از ترکیب صدها کانال طیفی در صحنه‌ای واحد، کاربرد گسترده‌ای دارند. طبقه‌بندی این تصاویر، به‌دلیل ماهیت پیچیدهٔ صحنهٔ تصویر (مثلاً داده‌های بسیار، پیکسل‌های مختلط و نمونه‌های آموزشی محدود)، مسئله‌ای پرچالش به‌شمار می‌آید و به‌همین دلیل تلاش‌های بسیاری برای پرداختن به آن انجام شده است. مجموعه داده‌های مورد استفاده، در این تحقیق، داده‌های جامع از زمین‌های پاینس در امریکا است که در ژوئن سال ۱۹۹۲ جمع‌آوری شده‌اند و شانزده کلاس را شامل می‌شوند که اغلب مربوط به پوشش زمینی‌اند (Ghassemian et al., 2003). در شکل ۱، این تصویر به‌همراه نقشهٔ کاربری نشان داده شده است.

این مجموعه داده، به‌منزلهٔ داده‌های پایه، برای ارائهٔ مدل‌های طبقه‌بندی در بیشتر تحقیقات بررسی و استفاده شده‌اند. داده‌های این مجموعه این امکان را فراهم می‌کند که مقایسات جامعی از رویکردها، در تحقیق پیش رو، انجام شود. ابعاد این تصاویر ابرطیفی ۲۰۰×۱۴۵×۱۴۵ است و اطلاعات هریک از دسته‌ها در جدول ۱ و پوشش زمینی در شکل ۷ آمده است.



(الف)



(ب)

شکل ۱. ترکیب رنگی تصویر Indian Pines (الف)؛ نقشهٔ کاربری تصویر Indian Pines (ب)

1. Xtreme Gradient Boosting

Archive of SID.ir

Archive of SID.ir

به‌منظور جلوگیری از ایجاد مشکل در تحلیل‌های عددی، مقادیر باندها نرمال‌سازی شد. برای نرمال‌سازی داده‌ها، از دو روش بهره گرفته شد؛ در روش تمامی باندها، مقدار درجهٔ خاکستری هر پیکسل در هر باند بر جمع مقادیر درجات خاکستری آن پیکسل در تمامی باندها تقسیم می‌شود و در روش تک‌باند نیز، مقدار درجهٔ خاکستری هر پیکسل در هر باند بر جمع مقادیر درجات خاکستری تمامی پیکسل‌های آن باند تقسیم می‌شود. تأثیر روش نرمالیزه‌کردن در نتایج طبقه‌بندی با استفاده از هستهٔ گاوسین بررسی و نشان داده شد، با هر دو روش، نتایج مشابهی به‌دست می‌آید؛ ازاین‌رو در این پژوهش، روش تمامی باندها استفاده شد.

۳- مبانی نظری

۳-۱- رویکردهای مبتنی‌بر یادگیری عمیق

در چند دههٔ اخیر، با افزایش قدرت پردازنده‌های گرافیکی و دردسترس‌بودن مجموعه‌های حجیم داده‌ها، استفاده از تکنیک‌های یادگیری عمیق، در بسیاری از شاخه‌های مهندسی و کاربردی، مطرح شده است. به‌موازات‌آن حوزه‌های پردازش تصویر نیز تحت تأثیر قرار گرفته‌اند؛ به‌گونه‌ای که در بسیاری از کاربردهای پردازش تصویر، یادگیری عمیق کارآیی بیشتری از تکنیک‌های یادگیری ماشین[1] سنتی داشته است (Xu et al., 2018; Li, 2018). تفاوت ساختار شبکه‌های عصبی کانولوشنی که یکی از تکنیک‌های یادگیری عمیق محسوب می‌شوند، با شبکه‌های عصبی مصنوعی، محدودیت‌های اعمال‌شده در ارتباطات بین نورون‌ها در شبکه‌های کانولوشنی است. اگرچه این محدودیت‌ها عمومیت شبکه‌های کانولوشن را در مواجهه با داده‌های متنوع کاهش می‌دهد، ساختار این شبکه‌ها با توجه به قیود خاص آن، کاملاً منطبق و مناسب با ساختار تصاویر است. همچنین در شبکه‌های عصبی کپسول، با توجه به اینکه خروجی شبکه بردار است، می‌توان ارتباط بین ویژگی‌های استخراجی از شبکه‌های کانولوشنی را درنظر گرفت (Maggiori et al., 2016). براین‌اساس، پژوهش حاضر بر فرایندهای تولید ویژگی‌های عمیق و طبقه‌بندی برمبنای آن متمرکز شده است؛ در ادامه، جزئیات روش‌های به‌کاررفته در هر مرحله را بیان می‌کنیم.

۳-۲- شبکه‌های عصبی کانولوشنی

شبکه‌های عصبی کانولوشنی نوعی از شبکه‌های رو به جلو هستند که اولین بار، برای تشخیص تصویر به‌کار رفتند. این شبکه‌ها از مهم‌ترین و پرکاربردترین

جدول ۱. پوشش زمینی کلاس های تصویر

شمارهٔ کلاس	تعداد پیکسل‌های کلاس در تصویر	پوشش زمینی	شمارهٔ کلاس	تعداد پیکسل‌های کلاس در تصویر	پوشش زمینی
۱	۵۴	یونجه	۹	۲۰	جو دوسر
۲	۱۴۳۴	ذرت بدون شخم	۱۰	۹۶۸	سویای بدون شخم
۳	۸۳۴	ذرت کم‌شخم	۱۱	۲۴۶۸	سویای کم‌شخم
۴	۲۳۴	ذرت	۱۲	۶۱۴	سویای شخم‌زده
۵	۴۹۷	سبزه/ چمن‌زار	۱۳	۲۱۲	گندم
۶	۷۴۷	سبزه/ درختان	۱۴	۱۲۹۴	بیشه
۷	۲۶	سبزه/ چمن‌زار کوتاه	۱۵	۳۸۰	ساختمان/ سبزه
۸	۴۸۹	کاه و خاشاک	۱۶	۹۵	سنگ/ برج

(بیشترین داده‌ها متعلق به کلاس‌های ۱۱، ۲ و ۱۴ است)

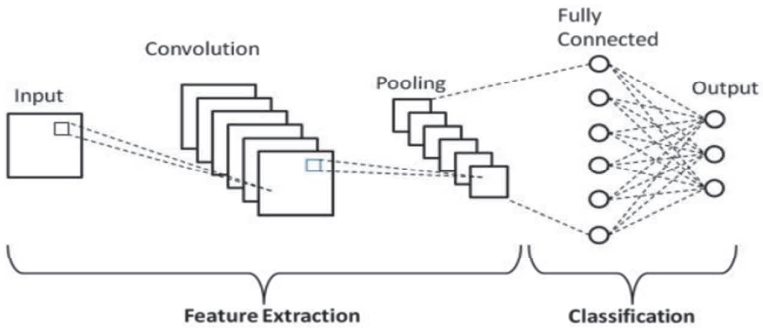
1. Machine Learning

Archive of SID.ir

Archive of SID.ir

روش‌های یادگیری عمیق، به‌ویژه درزَمینهٔ استخراج ویژگی و طبقه‌بندی تصاویر سنجش از دور، محسوب می‌شوند (LeCun et al., 2015). شبکه‌های عصبی کانولوشنی لایه‌های متعددی دارند که هریک از لایه‌ها یک سطح از ویژگی را نمایندگی می‌کنند. این سطوح متفاوت نمایندگی با کنار هم قرارگرفتن ساختارهای ساده و غیرخطی به‌دست می‌آیند و هریک از ساختارها ویژگی‌ها را به سطح انتزاع بالاتری ارتقا خواهد داد. شبکهٔ کانولوشنی چهار مشخصهٔ اصلی دارند: ارتباطات محلی، اشتراک وزن، لایهٔ ادغام و مشخصهٔ عمق که این مشخصات آن را از سایر روش‌ها متمایز می‌سازد. ساختار شبکه‌های کانولوشنی بخش‌های متوالی زیر را شامل می‌شود (Bengio & Ian, 2015):

- **لایهٔ کانولوشنی:** این لایه، به‌منظور استخراج ویژگی‌های ضمنی از ماتریس ورودی، از فیلترهای گوناگونی استفاده می‌کند؛
- **لایهٔ تجمیع:** این‌گونه معمول است که به‌صورت دوره‌ای، پس‌از هر لایهٔ کانولوشن، برای کاهش اندازهٔ ویژگی‌های میانی از یک لایهٔ تجمیع استفاده شود. کاهش حجم پارامترها سبب کاهش محاسبات شبکه می‌شود و ازاین‌رو می‌توان بیش‌برازش[1] را کنترل کرد؛
- **واحد اصلاح‌شدهٔ خطی (ReLu)**[2]: مزیت اصلی استفاده از این تابع بر سایر توابع این است که هم‌زمان تمامی نورون‌ها را فعال نمی‌کند؛ بدین‌معنی که در یک زمان، فقط نورون‌های اندکی فعال می‌شوند و این باعث می‌شود شبکه به‌صورت تنک دربیاید و ویژگی‌های مهمی را فرابگیرد؛
- **لایهٔ Dropout**: اصطلاح «Dropout» به واحدهای حذف‌شده (پنهان و آشکار) در شبکهٔ عصبی اشاره دارد. این تکنیک را اولین بار جفری هینتون[3] به‌کار برد (Bengio & Ian, 2015). Dropout درواقع از تکنیک‌های Regularization تصادفی است که طی هر انتشار رو به جلوی واحدها در لایه‌های پنهان یا آشکار، به‌صورت تصادفی یا موقت، مطابق یک بردار از متغیرهای تصادفی برنولی بیرون انداخته می‌شوند.
- **لایهٔ تماماً متصل**: پس‌از آخرین لایهٔ تجمیع، لایهٔ تماماً متصل[4] وجود دارد که ویژگی‌های میانی دوبعدی را به بردارهای یک‌بعدی، به‌منظور ادامهٔ فرایند یادگیری، نگاشت می‌کند. ازسویی، می‌توان گفت این لایه وظیفهٔ محاسبهٔ امتیاز دسته‌های طبقه‌ها را برعهده دارد. لایهٔ تماماً متصل به ما این امکان را می‌دهد که خروجی شبکه را در قالب برداری با اندازهٔ ثابت ارائه دهیم.



شکل ۲. معماری شبکه‌های کانولوشنی
منبع: LeCun et al., 2015

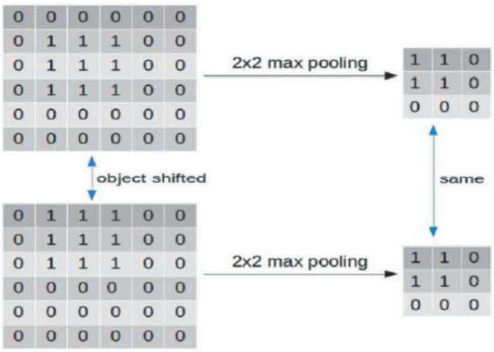
1. Overfitting
2. Rectified Linear Unit
3. Geoffrey Everest Hinton
4. Fully Connected Layer

Archive of SID.ir

Archive of SID.ir

۱-۲-۳- مزایای شبکهٔ کانولوشنی

مهم‌ترین مزیت استفاده از شبکهٔ عصبی کانولوشن توانایی آن در استخراج خودکارِ ویژگی‌های تصاویر، با استفاده از مفهوم یادگیری عمیق است. همچنین استفاده از لایهٔ pooling در شبکه‌های کانولوشنی باعث کاهش حجم محاسبات می‌شود و تضمین می‌کند، حتی زمانی‌که ویژگی‌های استخراجی دچار تغییرات و چرخش‌های کوچکی شوند، خروجی شبکه همچنان یکسان بماند. این ویژگی در طبقه‌بندی اهمیت بسیاری دارد. استفاده از max pooling نیز باعث می‌شود، با تغییرات اندک و جابه‌جایی بسیار کم در تصویر، شبکه همچنان قادر به شناسایی شیء مورد نظر باشد (شکل ۳).



شکل ۳. مزیت استفاده از لایهٔ pooling در شبکه‌های کانولوشنی
منبع: Bengio & Ian, 2015

۲-۲-۳- معایب شبکهٔ کانولوشنی

یکی از کاربردهای شبکه‌های کانولوشنی استخراج ویژگی‌ها برای طبقه‌بندی تصاویر است اما، زمانی‌که نمونهٔ معرفی‌شده به مدل آموزش‌دیدهٔ CNN تغییراتی داشته باشد، شبکه در تشخیص این تغییرات دچار مشکل می‌شود. دلیل این مشکل استفاده از لایهٔ Max-pooling است که فضای ارتباطی بین ویژگی‌ها را از بین می‌برد. در شکل ۴، تصویری از چهرهٔ درهم‌ریخته و غیرواقعی ارائه شده است که شبکه‌های CNN به‌اشتباه، آن را چهره تشخیص می‌دهند زیرا ارتباط بین ویژگی‌های استخراجی و ابعاد ویژگی‌ها را درنظر نمی‌گیرند. برای اینکه شبکه دچار اشتباه نشود، باید شبکهٔ کانولوشنی با تصاویری از چرخش‌های متفاوت آموزش ببینند؛ این کار نیازمند داده‌های آموزشی بسیار زیادی است.

با توجه به معایب بیان‌شده، استفاده از شبکهٔ کپسول برای رفع معایب شبکه‌های کانولوشنی پیشنهاد می‌شود.

۳-۳- شبکهٔ کپسول

شبکه‌های عصبی کپسول روش کاملاً جدیدی از معماری یادگیری عمیق است که سعی دارد بر محدودیت‌ها و اشکالات شبکهٔ کانولوشنی غلبه کند (Xue et al., 2015). خروجی‌های شبکهٔ عصبی کپسول در مجموعه‌ای از کپسول‌ها، به‌صورت بردار ذخیره



شکل ۴. تصویر چهره‌ای که نحوهٔ قرارگیری و ابعاد اجزای آن تغییر یافته است

می‌شوند و کپسول‌ها محاسبات پیچیده‌ای را روی این ویژگی‌ها انجام می‌دهند؛ سپس نتایج این محاسبات را به بردار کوچکی که حاوی اطلاعات بسیار مفیدی است، نگاشت می‌کنند. ازاین‌رو می‌توان ارتباط بین ویژگی‌های استخراجی شبکه را ازطریق اندازه و جهت بردار، حفظ کرد و هریک از المان‌های بردار خروجی نشان‌دهندهٔ توصیفی خاص از تصویر است (Candemir et al., 2014). یکی از مهم‌ترین تفاوت‌های شبکه‌های کپسول با شبکه‌های کانولوشنی در شکل ویژگی‌های استخراج‌شده است که به‌صورت برداری با ویژگی اندازه و جهت است؛ درحالی که در شبکه‌های کانولوشنی و

Archive of SID.ir

Archive of SID.ir

شبکه‌های تماماً متصل، خروجی‌ها به‌صورت عددی (اسکالر) هستند. شبکه‌های کپسول، مطابق شکل ۴، دارای لایه‌های Conv1d، PrimaryCaps و DigitCaps هستند (Xue et al., 2015; Candemir et al., 2014). با توجه به اینکه ویژگی‌ها به‌صورت بردار ذخیره شده‌اند، با تغییر آنها در جای‌جای تصویر، اندازهٔ بردار حفظ می‌شود اما جهت آن تغییر می‌کند و درنتیجه، برای شناسایی شیء با چرخش‌های متفاوت، به نمونه‌های آموزشی بیشتری نیاز نخواهد بود.

۱-۳-۳- اجزای شبکهٔ کپسول پیشنهادی و محاسبهٔ خروجی

شبکهٔ کپسول طراحی‌شده، در این پژوهش، بسیار شبیه به شبکهٔ کپسولی است که هینتون آن را توصیف کرده (Candemir et al., 2014)؛ با این تفاوت که تغییراتی روی آن اعمال شده است. لایه‌های شبکهٔ طراحی‌شده بدین‌قرار است:

- **لایهٔ کانولوشن**

این لایه دارای ۲۵۶ فیلتر کانولوشنی به‌اندازهٔ ۹×۹ و گام حرکتی ۲ است که روی آن، تابع فعال‌ساز «ReLu» اعمال شده است. یکی از تغییرات اعمال‌شده در این مرحله، در قیاس با شبکهٔ هینتون، تغییر مقدار گام حرکتی از ۱ به ۲ است زیرا مدل اصلی روی مجموعه داده‌های mnist استفاده شده بود و تصاویر mnist به‌صورت ۲۸×۲۸ است. با استفاده از گام حرکتی ۲، مقدار پارامترهایی که مدل در لایهٔ بعد دریافت می‌کند، بسیار کاهش می‌یابد.

- **لایهٔ PrimeCaps**

مطابق نسخهٔ اولیه، این لایه دارای کپسول‌هایی به‌اندازهٔ ۸، ۳۲ فیلتر ۹×۹ و گام حرکتی ۲ است.

- **DigitCaps**

نسخهٔ اولیهٔ این لایه دَه کپسول شانزده‌بعدی است که با توجه به مسئلهٔ تحقیق حاضر، این مقدار به دو کپسول شانزده‌بعدی تغییر یافت. در این لایه، مقدار الگوریتم مسیریابی نیز، همانند نسخهٔ اولیه، با مقدار ۳ تنظیم شد.

- **لایه‌های تماماً متصل**

برخلاف پژوهش هینتون که در آن لایه‌های تماماً متصل درنظر گرفته نشده است، در این مطالعه، لایه‌های مورد اشاره منظور شده‌اند و در آن، می‌توان از چندین لایه استفاده کرد. درواقع، این لایه سعی در فراگرفتن ویژگی‌های استخراجی از لایهٔ DigitCaps دارد. در این تحقیق، آخرین لایهٔ تماماً متصل به‌اندازهٔ طبقه‌های مد نظر این مطالعه (شانزده‌تا) است.

شکل ۵ ساختار کلی رویکرد شبکه‌های کپسول را نشان می‌دهد.

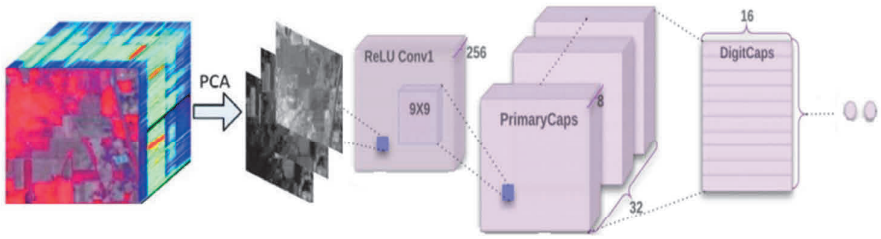
برای یک کپسول، مطابق شکل ۶، ورودی شبکه بردار u_i و خروجی آن بردار v_j است و پارامترها از یک لایه به لایهٔ دیگر منتقل می‌شوند.

برای ارسال کپسول از لایه‌ای به لایهٔ بعدی، به فرایند خاصی نیاز است که در ادامه، به آن می‌پردازیم.

برای این کار، یک ماتریس تبدیل w_i روی خروجی کپسول u_i لایهٔ پیشین اعمال می‌شود و سپس وزن جمعی s_j براساس وزن‌های c_{ij}، مطابق رابطه (۱)، محاسبه می‌شود (Candemir et al., 2014).

رابطه (۱)

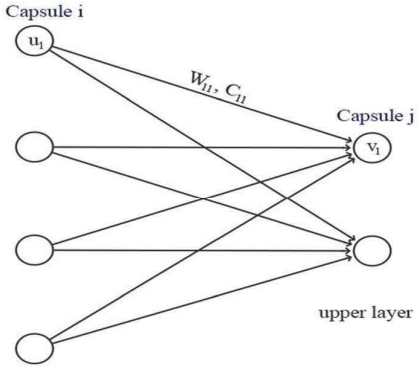
$$\widehat{U_{j|i}} = w_{ij} U_j$$
$$s_j = \sum c_{ij} . \widehat{U_{j|i}}$$



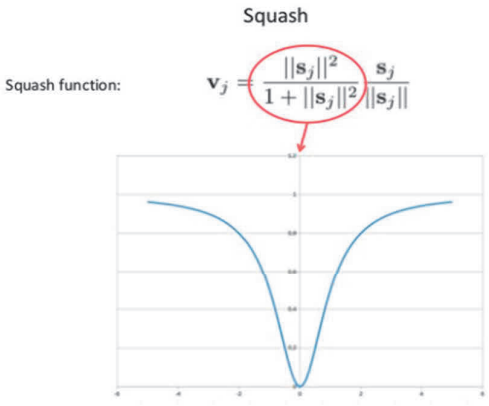
شکل ۵. شمای کلی از رویکرد شبکه‌های کپسول

Archive of SID.ir

Archive of SID.ir



شکل ۶. ارسال پارامترها از یک لایه به کپسول‌های لایهٔ بعدی



شکل ۷. شکل تابع Squash

منبع: Candemir et al., 2014

در این رابطه، c_{ij}ها ضرایب اتصالی‌اند که با استفاده از مسیریابی پویا[1] محاسبه می‌شوند و $\sum c_{ij}$ برای جمع آنها درنظر گرفته شده است. به‌طور مفهومی، c_{ij} معیاری برای اندازه‌گیری میزان احتمال فعال‌سازی کپسول j با کپسول i است. در شبکه‌های کپسول، به‌جای اعمال ReLu، تابع squash را روی خروجی s_j اعمال می‌کنند. این تابع بردارهای کوچک را به ۰ و بردارهای بزرگ را به بردارهای واحدی کاهش می‌دهد و به‌همین‌دلیل خروجی یک کپسول v_j، مطابق رابطه (۲)، طولی بین [0-1] دارد. در شکل ۷، تابع squash ارائه شده و فرایند اعمال squash روی بردار ورودی s_j نیز در ادامه، آورده شده است (Candemir et al., 2014).

رابطه (۲)

$$v_j = \frac{||s_j||^2}{1+||s_j||^2} \cdot \frac{s_j}{||s_j||}$$

$$v_j \approx ||s_j|| s_j \quad for\ small\ s_j$$

$$v_j \approx \frac{s_j}{||s_j||} \quad for\ larg\ s_j$$

لایهٔ آخر تماماً متصل، در مقایسه با رویکردهای سنتی یادگیری ماشین مانند SVM و XGBoost، دقت کمتر و میزان محاسبات بیشتری دارد؛ ازاین‌رو در رویکرد پیشنهادی، مدل XGBoost به‌منزلهٔ جایگزین این لایه‌ها به‌کار رفت.

۴-۳- مدل درخت تقویتی

مدل درخت تقویتی روشی بسیار مؤثر و پرکاربرد در یادگیری ماشین است که چن و گواستارین[2] (۲۰۱۶) آن را پیشنهاد کرده‌اند. این شبکه نوعی تکنیک اجرایی کارآمد برای افزایش صحت و عملکرد ماشین افزایش گرادیان (GBM)[3] و به‌ویژه، افزایش دقت طبقه‌بندی درخت‌های رگرسیون است. همچنین یکی از محبوب‌ترین و کارآمدترین عملی‌سازی و اجراهای الگوریتمGradient Boosted Trees و روش یادگیری نظارت‌شده‌ای مبتنی‌بر تقریب عملکرد با بهینه‌سازی توابع خاص است (Paoletti et al., 2019). Boosting در ML از تکنیک‌های یادگیری متوالی برای تبدیل فرضیه‌ای ضعیف به فرضیه‌ای قوی، به‌منظور افزایش دقت مدل به‌شمار می‌رود (Chen & Guestrin, 2016).

ساختار مدل پیشنهادی، پس‌از حذف لایهٔ تماماً متصل و جایگزینی آن با XGBoost، در شکل ۸ نشان داده شده است.

---

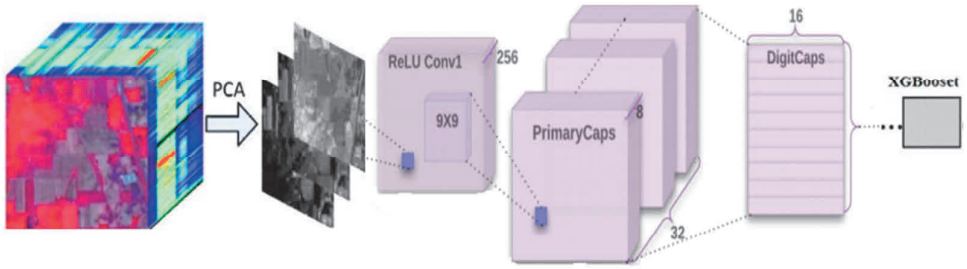
۱. در شبکه‌های کپسول، برای محاسبهٔ خروجی کپسول، از مسیریابی پویایی با محاسبهٔ ضریب اتصالی به‌صورت تکراری استفاده می‌کنند. این فرایند مشابه الگوریتم پس‌انتشار در شبکه‌های عصبی مصنوعی است؛ با این تفاوت که عمل مسیریابی به‌تعداد محدود انجام می‌شود (به‌صورت پیش‌فرض، سه بار).

2. Chen & Guestrin

3. Gradient Boosting Machine

Archive of SID.ir

Archive of SID.ir



شکل ۸. شمای کلی از رویکرد پیشنهادی تلفیق شبکه‌های کپسول و درخت تصمیم تقویتی

در این پژوهش، به‌منظور بررسی صحت روش پیشنهادی، نتایج با سایر روش‌های منابع پیشین نیز بررسی شد؛ در ادامه، توضیح مختصر روش‌ها خواهد آمد.

• رویکرد ماشین بردار پشتیبان (SVM)

ماشین بردار پشتیبان نوعی از طبقه‌بندی‌کننده‌های آماری است که عملکرد کلی آن بیشتر بر الگوهای طبقه‌بندی و تخمین رگرسیون غیرخطی تمرکز دارد (Landgrebe, 2002). ماشین‌های بردار پشتیبان، ازطریق انتقال فضای ورودی به فضایی با بعد بالاتر، می‌توانند مسائل غیرخطی را حل کنند؛ این کار ازطریق توابع کرنل امکان‌پذیر است. کرنل‌های متفاوت بسیاری همچون کرنل گوسی، کرنل خطی، کرنل چندجمله‌ای و کرنل تانژانت هایپربولیک وجود دارند. در پژوهش انجام‌شده، کرنل گوسی به‌منظور طبقه‌بندی تصاویر ابرطیفی به‌کار رفته است.

• رویکرد RF-200

الگوریتم جنگل تصادفی (Landgrebe, 2002) مبتنی‌بر دسته‌ای از درخت‌های تصمیم‌گیری است که برچسب نهایی هر پیکسل را براساس برآوردی از تصمیمات درخت‌ها و اغلب برمبنای رأی اکثریت، تعیین می‌کند و برای بسیاری از مجموعه داده‌ها، طبقه‌بندی با صحت بالایی را تضمین می‌کند.

• شبکه‌های عصبی بازگشتی RNN-LSTM

LSTM[1] نوعی مدل یا ساختار برای داده‌های ترتیبی است که در سال ۱۹۹۵ برای توسعهٔ شبکه‌های عصبی بازگشتی ظهور یافت (Keshavarz & Ghasemiyan, 2005). عبارت Long Term Memory به وزن‌های یادگرفته‌شده و Term Short Memory به حالت‌های درونی سلول‌ها اطلاق می‌شود. تغییر مهم و اصلی این ساختار، در قیاس با RNN، جایگزین‌کردن لایهٔ میانی RNN با یک بلوک است (Keshavarz & Ghasemiyan, 2005). مهم‌ترین ویژگی LSTMامکان یادگیری وابستگی بلندمدت است که ازطریق شبکه‌های عصبی بازگشتی، امکان‌پذیر نبود. برای پیش‌بینی گام زمانی بعدی، لازم است مقادیر وزن‌ها در شبکه به‌روزرسانی شوند که این کار مستلزم حفظ اطلاعات گام‌های زمانی ابتدایی است. در تحقیق مورد بررسی، شبکه‌های LSTM به‌منظور طبقه‌بندی تصویر ابرطیفی، اجرا شد. نتایج پژوهش در جدول ۳ آمده است (Keshavarz & Ghasemiyan, 2005).

• شبکه‌های عصبی بازگشتی RNN–GRU

معماری GRU[2] (Keshavarz & Ghasemiyan, 2005) به‌منظور حل کاستی‌های شبکهٔ عصبی بازگشتی قدیمی مانند مشکل محوشدگی گرادیان و نیز معایب معماری LSTM مطرح شده است. GRU اغلب نسخه‌ای تغییریافته از LSTM درنظر گرفته می‌شود زیرا هر دوی این معماری‌ها طراحی مشابهی دارند و در برخی موارد،

1. Long Short Term Memory
2. Gated Recurrent Unit

Archive of SID.ir

Archive of SID.ir

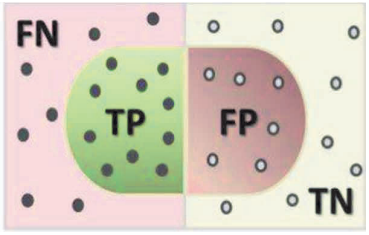
به‌صورت یکسان، نتایج عالی دربردارند. همان‌گونه که بیان شد، یکی از راه حل‌ها برای حل مشکل محوشدگی گرادیان در شبکهٔ عصبی سنتی، استفاده از GRU است. این نوع معماری از مفهوم‌هایی به‌نام دروازهٔ به‌روزرسانی[1] و دروازهٔ بازنشانی[2] استفاده می‌کند. این دو دروازه، درأصل، دو بردارند که با استفاده از آنها تصمیم گرفته می‌شود چه اطلاعاتی به خروجی منتقل شود و چه اطلاعاتی انتقال نیابد. نکتهٔ خاص دربارهٔ دروازه‌های مورد نظر این است که این دروازه‌ها را می‌توان طوری آموزش داد که اطلاعات مربوط به گام‌های زمانی بسیار قبل را حفظ کند؛ بدون آنکه حین گذر زمان (طی گام‌های زمانی گوناگون) دستخوش تغییر شوند.

• رویکرد RNN-GRU-PRetanh

در این رویکرد، شبکهٔ بازگشتی با بلوک‌های GRU و نیز تابع فعال‌ساز تانژانت هایپربولیک پارامتریک[3]، برای طبقه‌بندی تصویر ابرطیفی استفاده شده است (Keshavarz & Ghasemiyan, 2005). نتایج این رویکرد در جدول ۳ آمده است.

۵-۳- معیارهای ارزیابی

پیش‌از طرح و بررسی نتایج، معیارهای ارزیابی مورد استفاده را بررسی می‌کنیم. طبقه‌بندی دوکلاسه[4] معیارهای ارزیابی گوناگونی دارند که قبل‌از بررسی آنها، پارامترهای به‌کاررفته در این معیارها توضیح داده می‌شود. اگر طبقه‌بندی دوکلاسه روی مجموعه‌ای دادهٔ تست اعمال شود، نتایج زیر (شکل ۹) به‌دست می‌آید (Domingos, 2012).



شکل ۹. اعمال طبقه‌بندی دودویی روی یک مجموعه دادهٔ تست[5]

در شکل ۹، چهار عبارت TP[6]، FP[7]، TN[8] و FN[9] نمایش داده شده است. عبارت TP بیانگر تعداد نمونه‌های صحیح است که به‌درستی پیش‌بینی شده‌اند. عبارت FP گویای تعداد نمونه‌های نادرستی است که به‌درستی پیش‌بینی شده‌اند. TN بیانگر تعداد نمونه‌هایی نادرستی است که نادرست دسته‌بندی شده‌اند. عبارت FN نیز به تعداد نمونه‌های درستی اشاره دارد که نادرست دسته‌بندی شده‌اند.

در طبقه‌بندی دودویی، معیار Precision برای اشاره به نسبت بین نمونه‌های پیش‌بینی‌شدهٔ مثبت که دقیقاً مثبت‌اند، در قیاس با کل پیش‌بینی‌های مثبت انجام‌شده به‌کار می‌رود. این مسئله با رابطه (۳) نشان داده می‌شود (Ghamisi et al., 2017).

$$Precision = \frac{TP}{TP + FP}$$ رابطه (۳)

به‌طور کلی عبارت Precision به‌معنی توانایی بالا برای پیش‌بینی درست است.

معیار Recall بیانگر تعداد اسناد پیش‌بینی‌شدهٔ مثبت، به‌نسبت کل اسناد موجود، شامل پیش‌بینی مثبت و منفی است که طبق رابطه (۴) نشان داده می‌شود.

$$Recall = \frac{TP}{TP + FN}$$ رابطه (۴)

عبارت Recall به‌این‌معنی است که طبقه‌بندی مورد نظر توانسته بسیاری از اسناد را به‌درست طبقه‌بندی کند.

1. Update gate
2. Reset gate
3. Parametric Rectified Tanh (PRetanh)
4. Binary Classification
5. Apply Binary Classifications to a Test Data Set
6. True Positive
7. False Positive
8. True Negative
9. False Negative

Archive of SID.ir

Archive of SID.ir

معیار F-Measure که میانگین هارمونیک نیز نامیده شده، ترکیب دو معیار Precision و Recall است. این عبارت معمولاً F-Score متعادل نیز نامیده و به‌صورت رابطه (۵) نشان داده می‌شود (Ghamisi et al., 2017).

$$F_{score} = 2 \times \frac{Precision * Recall}{Precision + Recall} \qquad \text{رابطه (۵)}$$

Accuracy یکی دیگر از معیارهایی است که در طبقه‌بندی دودویی کاربرد دارد. برای محاسبهٔ این معیار، نسبت TP و TN به تعداد کل موارد پیش‌بینی‌شده درنظر گرفته می‌شود؛ بنابراین رابطه (۶) برای آن به‌دست می‌آید.

$$Accuracy = \frac{TP + FN}{TP + FP + FN + TN} \qquad \text{رابطه (۶)}$$

۶-۳- پارامترهای شبکه

در پژوهش پیش رو، اندازهٔ نمونه‌های آموزشی ۳۰×۵×۵ است؛ ۵×۵ به اندازهٔ دسته[1] اشاره دارد و ۳۰ نتیجهٔ اعمال الگوریتم کاهش بعد PCA روی باندهای تصویر ابرطیفی است. در بررسی‌های انجام‌شدهٔ ما، اندازهٔ ۳۰ بهینه‌ترین حالت برای کاهش بعد PCA است؛ دراین‌حالت تغییرات در معیارها بسیار ناچیز است و به‌طور کلی در افزایش بعد، بهبود حاصل نمی‌شود. درمجموع، اندازهٔ فیلترهای کانولوشنی در شبکه‌های CNN یا کپسول ممکن است ۳، ۵ و یا بیشتر باشد. در این تحقیق، از فیلتر ۹×۹ استفاده شده است. لایهٔ کانولوشنی، در ساختار رویکرد پیشنهادی، دارای ۲۵۶ فیلتر کانولوشنی به‌اندازهٔ ۹×۹ و گام حرکتی ۲ است که روی آن تابع فعال‌ساز ReLu اعمال شده است. در این ساختار رویکرد همچنین، کپسول‌هایی به‌اندازهٔ ۸ با ۳۲ فیلتر ۹×۹ و گام حرکتی ۲ به‌کار رفت و میزان Dropout، در تمامی لایه‌های شبکه، ۰/۵ درنظر گرفته شد. تمامی پارامترهای مدل با استفاده از Optuna[2] به‌دست آمده‌اند. Optuna کتابخانه‌ای متن باز برای جست‌وجوی پارامترهای بهینه در مدل‌های یادگیری عمیق است. ۸۰٪ داده‌های این تحقیق، به‌منزلهٔ داده‌های آموزشی، برای آموزش مدل و ۲۰٪ به‌صورت داده‌های آزمون ارزیابی مدل، استفاده خواهند شد.

۴- نتایج

در این بخش، نتایج حاصل از شبکهٔ کپسول پیشنهادی و شبکهٔ CNN با استفاده از معیارهای ارزیابی Precision، Recall و F1-score (Ghamisi et al., 2017) مقایسه می‌شود. رویکرد پیشنهادی نیز با پنج رویکرد مرسوم طبقه‌بندی مقایسه شده است که در ادامه، نتایج هریک از این مدل‌ها بیان می‌شود.

۴-۱- نتایج رویکرد مدل شبکهٔ کانولوشنی

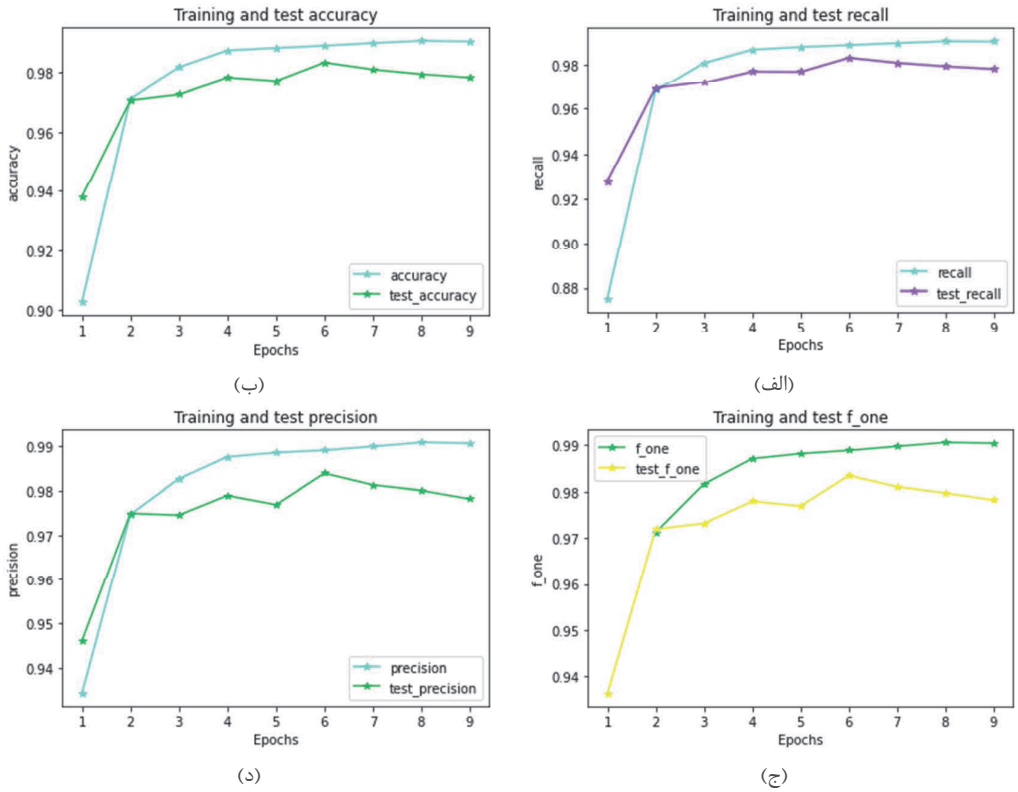
در این بخش، نتایج اجرای شبکهٔ CNN روی دادهٔ پاینس هند بیان شده است. نتایج در دو حالت، به‌ازای هر کلاس مجزا، به‌ازای تمامی حالت‌ها و برای مقایسه با سایر رویکردها گزارش داده شده است. برای تحلیل‌های بیشتر، نمودار خطا و متریک‌های اندازه‌گیری در شکل ۱۰ آورده شده است. جدول ۲ نیز نتیجهٔ اجرای شبکهٔ CNN روی مجموعهٔ داده‌های تست، بعدآز پنجاه بار تکرار مدل است که به‌ازای هر کلاس ارائه شده است. ستون Count تعداد نمونه به‌ازای هر کلاس را نشان می‌دهد.

در این نمودارها، محور X نشان‌دهندهٔ تعداد دفعات اجرای الگوریتم روی داده‌ها و محور Y گویای معیار به‌دست‌آمده، به‌ازای هر بار اجرای الگوریتم است. در شکل ۱۱ نیز، ماتریس خطای مرتبط با رویکرد اول ارائه شده است. جدول ۳ دقت نتایج به‌دست‌آمده با استفاده از ساختار اول و مقایسهٔ آن با نتایج دیگر پژوهش‌ها را درمورد این مجموعه از داده‌ها، بازگو می‌کند.

1. Patch Size
2. https://optuna.org/

Archive of SID.ir

Archive of SID.ir



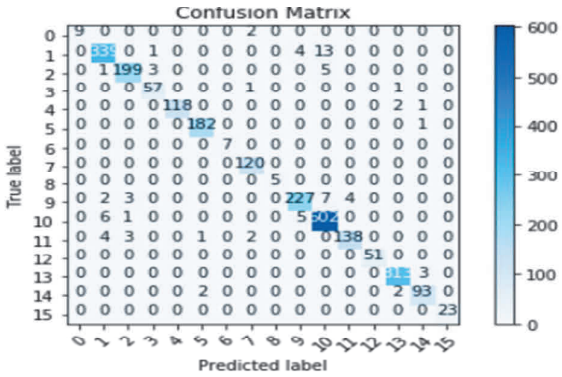
شکل ۱۰. نمودار Recall (الف)؛ نمودار Accuracy (ب)؛ نمودار Precision (ج)؛ نمودار F1 روی داده‌های آموزش و تست با رویکرد پیشنهادی اول (د)

جدول ۲. نتایج حاصل از ساختار پیشنهادی اول روی مجموعه داده‌های تست

شمارهٔ کلاس	معیارهای ارزیابی			Count
	Precision	Recall	F1-score	
۱	۱۰۰	۸۲	۹۰	۱۱
۲	۹۶	۹۵	۹۶	۳۵۷
۳	۹۷	۹۶	۹۶	۲۰۸
۴	۹۳	۹۷	۹۵	۵۹
۵	۱۰۰	۹۸	۹۹	۱۲۱
۶	۹۹	۱۰۰	۹۹	۱۸۳
۷	۱۰۰	۱۰۰	۱۰۰	۷
۸	۹۶	۱۰۰	۹۸	۱۲۰
۹	۱۰۰	۱۰۰	۱۰۰	۵
۱۰	۹۶	۹۴	۹۵	۲۴۳
۱۱	۹۶	۹۸	۹۷	۶۱۴
۱۲	۹۷	۹۳	۹۵	۱۴۸
۱۳	۹۸	۱۰۰	۱۰۰	۵۱
۱۴	۹۸	۹۹	۹۹	۳۱۶
۱۵	۹۵	۹۶	۹۵	۹۷
۱۶	۱۰۰	۱۰۰	۱۰۰	۲۳
میانگین کل	۹۶/۸۸	۹۶/۸۷	۹۶/۸۶	۲۵۶۳

Archive of SID.ir

Archive of SID.ir



شکل ۱۱. ماتریس خطای[1] رویکرد اول

جدول ۳. مقایسهٔ نتایج حاصل از ساختار پیشنهادی اول با نتایج موجود (مقادیر برحسب درصد است)

شمارهٔ کلاس	رویکرد اول	RNN-GRU-Pretanh	RNN-GRU-than	RNN-LSTM	SVM-RBF	RF-200
۱	۹۷/۵۶	۷۰/۵۹	۶۸/۹۳	۴۶/۰۳	۶۰/۷۷	۵۸/۸۴
۲	۹۷/۵۲	۷۰/۲۸	۴۰/۹۴	۶۱/۳۷	۷۷/۶۸	۵۸/۴۲
۳	۸۵/۳۳	۸۱/۵۲	۷۸/۸۰	۸۶/۹۶	۷۹/۳۵	۸۲/۶۱
۴	۸۷/۹۲	۹۰/۱۶	۸۷/۹۲	۸۷/۰۲	۹۱/۰۵	۸۵/۹۱
۵	۸۵/۲۲	۹۱/۹۷	۸۷/۵۲	۸۶/۶۶	۸۴/۳۶	۸۰/۴۹
۶	۹۷/۴۹	۹۶/۱۳	۹۷/۲۷	۷۹/۴۹	۹۲/۰۳	۹۴/۷۶
۷	۷۴/۶۲	۸۴/۷۵	۸۲/۷۹	۵۹/۶۹	۶۹/۶۱	۷۷/۳۴
۸	۶۷/۹۹	۵۹/۶۴	۵۰/۵۸	۶۴/۸۹	۵۹/۳۱	۵۹/۴۳
۹	۸۷/۵۸	۸۶/۱۷	۷۹/۴۳	۶۰/۴۶	۷۹/۶۱	۶۳/۴۸
۱۰	۹۸/۷۷	۹۹/۳۸	۹۸/۷۷	۹۸/۷۷	۹۷/۵۳	۵۹/۰۶
۱۱	۸۷/۶۲	۸۴/۹۷	۸۴/۷۳	۷۵/۳۲	۸۵/۲۱	۸۸/۲۶
۱۲	۷۲/۴۲	۷۷/۵۸	۶۱/۲۱	۷۱/۸۲	۶۳/۶۴	۵۴/۸۵
۱۳	۹۳/۳۳	۹۵/۵۶	۸۸/۸۹	۹۱/۱۱	۱۰۰	۹۷/۷۸
۱۴	۷۱/۷۹	۸۴/۶۲	۷۹/۴۹	۷۹/۴۹	۸۷/۱۸	۵۸/۹۷
۱۵	۹۰/۹۱	۹۰/۹۱	۹۰/۹۱	۹۰/۹۱	۹۰/۹۱	۸۱/۸۲
۱۶	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰

مقادیر ساختار اول، در مقایسه با رویکردهای قبلی، دقت شایان توجهی به‌دست آورده است. دقت این مدل، در بیشتر دامنه‌ها، کمتر از رویکرد RNN-GRU-PRetanh است و همچنین فقط توانسته روی داده‌های کلاس Buildings-grass-trees دقت برابری کسب کند. این درحالی است که در سایر طبقه‌ها، میزان دقت نسبتاً کمتر بوده است. این مدل به‌دلیل داشتن عمق و تعداد فیلتر کمتر، در طبقه‌بندی تصاویر، دقت کمتری داشته ازاین‌رو در رویکرد پیشنهادی دوم، سعی شده است با استفاده از شبکه‌های کپسول، تاحد بسیاری از این مشکلات کاسته شود. در هر دو رویکرد پیشنهادی، برای کنترل Overffiting از Early Stoping استفاده شده است. این سیاست، طی دو دور اجرایی، بررسی می‌کند که اگر مقدار Loss تغییری نکرده باشد، یادگیری را متوقف کند. برای یادگیری پارامترهای هر دو مدل نیز، Optuna به‌کار رفت و بهینه‌ترین پارامترها برای طبقه‌بندی انتخاب شد.

Archive of SID.ir

Archive of SID.ir

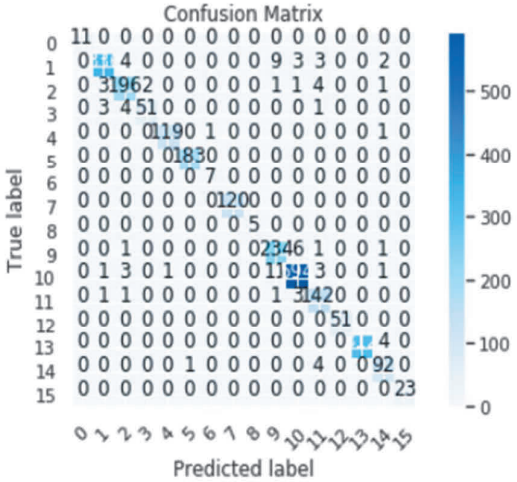
۲-۴- نتایج رویکرد پیشنهادی دوم شبکه‌های کپسول

در رویکرد پیشنهادی دوم نیز، از ۸۰٪ داده‌ها برای آموزش و از ۲۰٪ آنها برای تست استفاده شده است. مشابه رویکرد پیشنهادی اول، نتایج در دو حالت (به‌ازای هر کلاس مجزا، به‌ازای تمامی حالت‌ها) برای مقایسه با سایر رویکردها گزارش شده است. جدول ۴ نتایج خروجی از الگوریتم روی مجموعه داده‌های تست، پس‌از سی بار یادگیری شبکه را بیان می‌کند که به‌ازای هر کلاس، گزارش داده شده است. در شکل ۱۲، ماتریس خطای مرتبط با رویکرد دوم ارائه شده است.

نمودار خطا و سایر متریک‌های اندازه‌گیری در شکل ۱۳ آورده شده است.

جدول ۴. نتیجهٔ حاصل از رویکرد پیشنهادی دوم روی مجموعه داده‌های تست

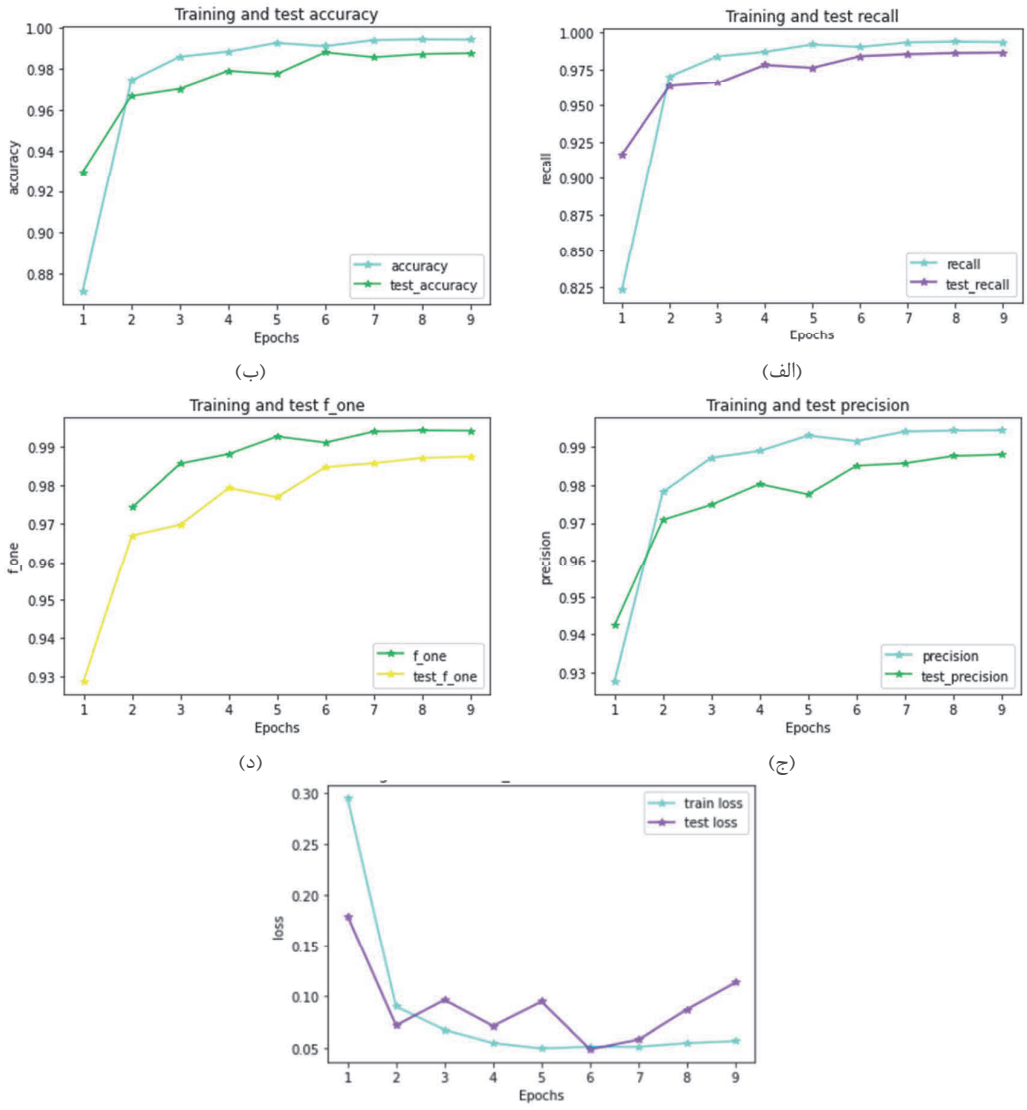
شمارهٔ کلاس	معیارهای ارزیابی			Count
	Precision	Recall	F1-score	
۱	۱۰۰	۱۰۰	۱۰۰	۱۱
۲	۹۸	۹۴	۹۶	۳۵۷
۳	۹۴	۹۴	۹۴	۲۰۸
۴	۹۶	۸۷	۹۱	۵۹
۵	۹۹	۹۸	۹۹	۱۲۱
۶	۹۹	۱۰۰	۱۰۰	۱۸۳
۷	۸۸	۱۰۰	۹۳	۷
۸	۱۰۰	۱۰۰	۱۰۰	۱۲۰
۹	۱۰۰	۱۰۰	۱۰۰	۵
۱۰	۹۲	۹۶	۹۴	۲۴۳
۱۱	۹۸	۹۷	۹۷	۶۱۴
۱۲	۹۰	۹۶	۹۳	۱۴۸
۱۳	۱۰۰	۱۰۰	۱۰۰	۵۱
۱۴	۱۰۰	۹۹	۹۹	۳۱۶
۱۵	۹۰	۹۵	۹۲	۹۷
۱۶	۱۰۰	۱۰۰	۱۰۰	۲۳
میانگین کل	۹۶/۶۸	۹۶/۶۰	۹۶/۶۲	۲۵۶۳



شکل ۱۲. ماتریس خطای رویکرد دوم

Archive of SID.ir

Archive of SID.ir



شکل ۱۳. نمودار Recall (الف)؛ نمودار Accuracy (ب)؛ نمودار Precision (ج)؛ F1 (د)؛ نمودار loss روی داده‌های آموزش و تست، با رویکرد پیشنهادی دوم (ه)

منحنی‌های حاصل، با رویکرد پیشنهادی دوم، کاملاً متفاوت از منحنی‌های حاصل از رویکرد پیشنهادی اول است. در این رویکرد، اختلاف بین دو منحنی آموزش و تست در تمامی حالات، کمتر شده است و این منحنی‌ها به مجانبی افقی میل کرده‌اند. این نکته نشان می‌دهد افزایش تعداد یادگیری شبکه، از مقداری معین، تأثیری در هیچ‌یک از معیارهای ارزیابی ندارد و هر چهار معیار، حدوداً پس‌از هفت مرحله یادگیری شبکه، بدون تغییر باقی می‌مانند. در جدول ۵، مقایسهٔ دقت‌های به‌دست‌آمده ازطریق رویکرد پیشنهادی دوم، با نتایج حاصل از دیگر پژوهش‌ها دربارهٔ این مجموعه از داده‌ها، آمده است.

Archive of SID.ir

Archive of SID.ir

جدول ۵. مقایسهٔ نتایج حاصل از ساختار پیشنهادی دوم، با نتایج سایر پژوهشگران

شمارهٔ کلاس	رویکرد اول	رویکرد دوم	RNN-GRU-Pretanh	RNN-GRU-than	RNN-LSTM	SVM-RBF	RF-2000
۱	۹۷/۵۶	۹۵/۷۱	۷۰/۵۹	۶۸/۹۳	۶۴/۰۳	۶۰/۷۷	۵۴/۸۴
۲	۹۷/۵۲	۸۲/۷۷	۷۰/۲۸	۴۰/۹۴	۶۱/۳۷	۷۷/۶۸	۵۸/۴۲
۳	۸۵/۳۳	۸۶/۵۲	۸۱/۵۲	۷۸/۸۰	۸۶/۹۶	۷۹/۳۵	۸۲/۶۱
۴	۸۷/۹۲	۹۱/۹۳	۹۰/۱۶	۸۷/۹۲	۸۷/۰۲	۹۱/۰۵	۸۵/۹۱
۵	۸۵/۲۲	۹۲/۹۷	۹۱/۹۷	۸۷/۵۲	۸۶/۶۶	۸۴/۳۶	۸۰/۴۹
۶	۹۷/۴۹	۹۸/۱۲	۹۶/۱۳	۹۷/۲۷	۷۹/۴۹	۹۲/۰۳	۹۴/۷۶
۷	۷۴/۶۲	۸۴/۷۰	۸۴/۷۵	۸۲/۷۹	۵۹/۶۹	۶۹/۶۱	۷۷/۳۴
۸	۶۷/۹۹	۶۷/۶۴	۵۹/۶۴	۵۰/۵۸	۶۴/۸۹	۵۹/۳۱	۵۹/۴۳
۹	۸۷/۵۸	۷۶/۹۰	۸۶/۱۷	۷۹/۴۳	۶۰/۴۶	۷۹/۶۱	۶۳/۴۸
۱۰	۹۸/۷۷	۹۹/۶۰	۹۹/۳۸	۹۸/۷۷	۹۸/۷۷	۹۷/۵۳	۵۹/۰۶
۱۱	۸۷/۶۲	۸۵/۹۰	۸۴/۹۷	۸۴/۷۳	۷۵/۳۲	۸۵/۲۱	۸۸/۲۶
۱۲	۷۲/۴۲	۷۸/۵۹	۷۷/۵۸	۶۱/۲۱	۷۱/۸۲	۶۳/۶۴	۵۴/۸۵
۱۳	۹۳/۳۳	۹۶/۵۰	۹۵/۵۶	۸۸/۸۹	۹۱/۱۱	۱۰۰	۹۷/۷۸
۱۴	۷۱/۷۹	۷۹/۶۲	۸۴/۶۲	۷۹/۴۹	۷۹/۴۹	۷۸/۱۸	۵۸/۹۷
۱۵	۹۰/۹۱	۹۱/۹۱	۹۱/۹۱	۹۰/۹۱	۹۰/۹۱	۹۰/۹۱	۸۱/۸۲
۱۶	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰

نتایج رویکرد پیشنهادی دوم، در بیشتر کلاس‌ها، بهتر از رویکرد اول بوده و این ساختار توانسته است دقت بیشتر از RNN-GRU-PRetanh به‌دست آورد؛ به این دلیل که رویکرد دوم شبکه‌های کپسول، با استخراج مطلوب ویژگی‌ها، توانسته است مهندسی ویژگی خودکار مناسبی برای مدل استاندارد XGBoost ایجاد کند. همچنین با استفاده از رویکرد دوم در دامنه‌های Alfalfa، Corn-min، Grass-pasture، Grass-trees، Soybean-notill، Soybean-clean، Wheat و Buildings-grass-trees دقت بیشتری حاصل شده است. ساختار دوم در سایر کلاس‌ها نیز دقت مورد قبولی داشته و البته علت اندک‌بودن دقت این ساختار، روی این طبقه‌ها، کم‌بودن تعداد نمونه‌ها در آنهاست که درنتیجهٔ‌آن، مدل به‌خوبی نتوانسته است این داده را آموزش ببیند و یادگیری را کامل کند.

۵- بحث و نتیجه‌گیری

طبقه‌بندی تصاویر ماهواره‌ای همواره با چالش‌های بسیاری روبه‌رو بوده است. با پیشرفت و گسترش الگوریتم‌های طبقه‌بندی و افزایش قدرت تفکیک طیفی تصاویر، با وجود افزایش کلی دقت، چالش‌های طبقه‌بندی این تصاویر نیز بیشتر شده است.

یکی از مهم‌ترین مشکلات طبقه‌بندی این تصاویر، در مدل‌های یادگیری ماشین، انتخاب ویژگی‌هاست زیرا این مدل‌ها نیاز به انتخاب ویژگی به‌صورت دستی دارند و هرچه ویژگی‌های مناسب‌تری استخراج شود، این مدل‌ها می‌توانند طبقه‌بندی بهتری انجام دهند. ازاین‌رو با یادگیری عمیق، به‌منزلهٔ مدل‌هایی که می‌توانند ویژگی‌ها را به‌صورت خودکار انتخاب کنند، جایگزین مناسبی برای طبقه‌بندی این تصاویر شمرده می‌شوند. در این نوع یادگیری، ویژگی‌ها براساس فیلترهایی استخراج می‌شود که کاربر هیچ دخالتی در آن ندارد. برای افزایش دقت طبقه‌بندی این تصاویر، تکنیک‌های یادگیری عمیق شبکه‌های کانولوشنی روی این نوع تصاویر اعمال می‌شود. در پژوهش حاضر، از روش یادگیری عمیق شبکه‌های عصبی کانولوشنی و نیز شبکه‌های کپسول، برای افزایش دقت طبقه‌بندی تصاویر ابرطیفی استفاده شد. ابعاد تصویر ورودی به

Archive of SID.ir

Archive of SID.ir

شبکه از عوامل تأثیرگذار در عملکرد استخراج ویژگی با استفاده از شبکه‌های کپسول است. در پژوهش حاضر، ابعاد ۳۰×۵×۵ درنظر گرفته شد. سه معیار Precision، Recall و F1-Score برای ارزیابی نتایج به‌کار رفت و نتایج، به‌صورت میانگین و انحراف معیار برای بیست بار اجرای شبکه، تهیه شد. از مجموع داده‌های درنظر گرفته‌شده، ۸۰٪ برای آموزش طبقه‌بندی و ۲۰٪ برای ارزیابی طبقه‌بندی استفاده شد. به‌منظور بهبود دقت طبقه‌بندی تصاویر نیز، دو رویکرد متفاوت درزمینهٔ استفاده از شبکه‌های عصبی کانولوشنی و شبکهٔ کپسول، در طبقه‌بندی تصاویر هوایی ابرطیفی به‌کار رفت. یکی از مهم‌ترین تفاوت‌های شبکه‌های کپسول با شبکه‌های کانولوشنی در ذخیرهٔ داده‌ها، در قالب یک بردار است. ویژگی بردار‌بودن ویژگی اندازه و جهت را نیز به‌همراه خواهد داشت و می‌توان ارتباط بین ویژگی‌ها را نیز استخراج کرد و به دقت بالاتری رسید. در رویکرد اول، از این شبکه‌ها در حالت یکپارچه برای استخراج ویژگی و در ادامه، طبقه‌بندی استفاده شد و در رویکرد دوم، ویژگی‌های استخراج‌شده ازطریق شبکه‌های کپسول، به‌منزلهٔ ورودی، به مدل درخت تقویتی XGBoost برای طبقه‌بندی معرفی شد. در رویکرد اول، از دو لایهٔ کانولوشنی و دو لایهٔ تماماً متصل استفاده شد که حصول دقت ۹۸٪ روی‌داده‌های آموزش و دقت ۹۶/۳۰٪ روی‌داده‌های تست را درپی داشت. در رویکرد دوم، از یک لایهٔ کانولوشنی، یک لایهٔ کپسول و XGboost استفاده شد که به دقت ۹۹٪ روی داده‌های آموزش و نیز دقت ۹۷/۵٪ روی داده‌های تست انجامید. رویکرد پیشنهادی اول فقط در یکی از طبقه‌ها توانسته است دقت برابر با رویکرد برتر در سایر پژوهش‌ها به‌دست آورد؛ درصورتی که رویکرد دوم، در بیشتر موارد، توانسته است به دقت بالاتر دست یابد.

۶- منابع

Akbari, D., Safari, A.R. & Homayouni, S., 2016, **Improving Spectral-Spatial Classification of Supercritical Images by Using Spatial Information to Select Symbols**, Scientific - Research Quarterly of Geographical Data (SEPEHR), 25(98), PP. 5-14.

Bengio, Y. & Ian, J., 2015, **Goodfellow, and Aaron Courville. "Deep Learning"**, Nature, 521.7553, PP. 436-444.

Bruzzone, L., Chi, M. & Marconconi, M., 2006, **A Novel Transductive SVM for Semisupervised Classification of Remote-Sensing Images**, IEEE Transactions on Geoscience and Remote Sensing, 44(11), PP. 3363-3373.

Candemir, S., Jaeger, S., Palaniappan, K., Musco, J.P., Singh, R.K., Xue, Z. & McDonald, C.J., 2014, **Lung Segmentation in Chest Radiographs Using Anatomical Atlases with Nonrigid Registration**, IEEE Transactions on Medical Imaging, 33(2), PP. 577-590.

Camps-Valls, G., Gomez-Chova, L., Munoz-Mari, J., Vila-Frances, J. & Calpe-Maravilla, J., 2006, **Composite Kernels for Hyperspectral Image Classification**, IEEE Geoscience and Remote Sensing Letters, 3(1), PP. 93-97.

Camps-Valls, G., Shervashidze, N. & Borgwardt, K.M., 2010, **Spatio-Spectral Remote Sensing Image Classification with Graph Kernels**, IEEE Geoscience and Remote Sensing Letters, 7(4), PP. 741-745.

Chen, T. & Guestrin, C., 2016, **Xgboost: A Scalable Tree Boosting System**, In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining.

Dehghani, H., 2006, **Classification of Distance Measurement Images with Large Dimensions and Limited Educational Examples**, Ph.D. Thesis, Tarbiyat Modarres University, Department of Electronic Engineering, Tehran, Iran.

Domingos, P., 2012, **A Few Useful Things to Know about Machine Learning**, Communications of the ACM.

Du, P., Tan, K., Zhang, W. & Yan, Zh., 2008, **ANN Classification of OMIS Hyperspectral Remotely Sensed Imagery: Experiments and Analysis**, Congress on Image and Signal Processing, IEEE.

Archive of SID.ir

Archive of SID.ir

Fauvel, M., Chanussot, J., Benediktsson, J.A. & Sveinsson, J., 2007, **Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles**, IEEE Transactions on Geoscience and Remote Sensing, 46(11), PP. 2012-2020.

Fukunaga, A. & Sber, G., 2008, **Providing an Optimal Method Based on Deep Learning for Spectral Classification Images with High Resolution Spatial Resolution in Semi-Urban Areas**, Journal of Geomatics Science and Technology, 9(2), PP. 151-170.

Gao, Y., Gao, F., Dong, J. & Li, H.C., 2020, **SAR Image Change Detection Based on Multiscale Capsule Network**, IEEE Geosci. Remote Sens. Lett., 18(3).

Ghaffari, O., Voldan-Zoj, M.J. & Mokhtarzadeh, M., 2016, **Selecting the Band in Order to Optimize the Spectral Separation of Supercritical Images**, The 1st National Conference on Geospatial Information Technology, K.N. Toosi University of Technology, Tehran, Iran.

Ghamisi, P., Plaza, J., Chen, Y., Li, J. & Plaza, A.J., 2017, **Advanced Spectral Classifiers for Hyperspectral Images**: **A Review**, IEEE Geoscience and Remote Sensing Magazine, 5(1), PP. 8-32.

Ghassemian, H., Keshavarz, A. & Landgrebe, D., 2003, **Hyper-Spectral Image Processing and Analyses**, Space Magazine, 1(3), PP. 32-41.

Gualtieri, J.A. & Chetri, S.R., 2000, **Support Vector Machines for Classification of Hyperspectral Data**, Proc. IGARSS, Honolulu, _HI, PP. 813-815.

Gualtieri, J.A. & Cromp, R.F., 1999, **Support Vector Machines for Hyperspectral Remote Sensing Classification**, Proceedings Vol. 3584, 27th AIPR Workshop: Advances in Computer-Assisted Recognition.

Gualtieri, J.A., Chetri, S.R., Cromp, R.F. & Johnson, L.F., 1999, **Support Vector Machines Classifiers as Applied to AVIRIS Data**, In Summaries 8th JPL Airborne Earth Sience Workshop, JPL Pub., 99-17, PP. 217-227.

Hu, W., Huang, Y., Wei, L., Zhang, F. & Li, H., 2015, **Deep Convolutional Neural Networks for Hyperspectral Image Classification**, J. Sens, 2015, P. 258619.

Jimenez, L. & Landgrebe, D.A., 1999, **Hyperspectral Data Analysis and Feature Reduction via Projection Pursuit**, IEEE Trans. On Geoscience and Remote Sensing, 37(6), PP. 2653-2667.

Kaewpijit, S., Moigne, J.L., Ghazawi, T.E., 2003, **Automatic Wavelet Spectral Analysis for Reduction of Hyperspectral Imagery**, IEEE Trans. on Geoscience and Remote Sensing, 41(4), PP. 863-871.

Keshavarz, A. & Ghasemiyan, H., 2005, **A High-Speed Vector Machine-Based Algorithm for Classifying Hayperspectral Images Using Spatial Correlation**, Iranian Journal of Electrical and Computer Engineering, 3(1), PP. 37-44.

Landgrebe, D.A., 2002, **Hyperspectral Image Data Analysis**, IEEE Signal Processing Magazine, 19(1), PP. 17-28.

LeCun, Y., Bengio, Y. & Hinton, G., 2015, **Deep Learning**, Nature, 521, PP. 436-444.

Lee, C. & Landgrebe, D.A., 1993, **Feature Extraction Based on Decision Boundaries**, IEEE Trans. on Pattern Analysis and Machine Intelligence, 15(4), PP. 388-400.

Li, H., 2018, **Deep Learning for Natural Language Processing: Advantages and Challenges**, National Science Review., 5(1), PP. 24-26.

Liyang, W., Yongyi, Y., Nishikawa, R.M., Wernick, M.N. & Edwards, A., 2005, **Relevance Vector Machine for Automatic Detection of Clustered Micro Calcifications**, IEEE Trans. Med. Imag., 24(10), PP. 1278-1285.

Luo, Y., Zou, J., Yao, C., Li, T. & Bai, G., 2018, **HSI-CNN: A Novel Convolution Neural Network for Hyperspectral Image**., arXiv 2018, arXiv:1802.10478.

Maggiori, E., Tarabalka, Y., Charpiat, G. & Alliez, P., 2016, **High-Resolution Semantic Labeling with Convolutional Neural Networks**, ArXiv: 1611.01962.

Archive of SID.ir

Archive of SID.ir

Melgani, F. & Bruzzone, L., 2004, **Classification of Hyperspectral Remote Sensing Images with Support Vector Machines**, IEEE Transactions on Geoscience and Remote Sensing, 42(8), PP.1778-1790.

Paoletti, M.E., Plaza, J. & Plaza, A., 2018, **A New Deep Convolutional Neural Network for Fast Hyperspectral Image Classification**, ISPRS Journal of Photogrammetry and Remote Sensing, 145, PP. 120-147. doi:10.1016/ j.isprsjprs.2017.11.021.

Paoletti, M.E., Haut, J.M., Fernandez-Beltran, R., Plaza, J., Plaza, A., Jun, L. & Pla, F., 2019, **Capsul Networks for Hyperspectral Image Classification**, IEEE Transactions on Geoscience and Remote Sensing, 57(4), PP. 2145-2160. doi:10.1109/TGRS.2018.2871782.

Sabour, S., Frosst, N. & Hinton, G.E., 2017, **Dynamic Routing between Capsules**, In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4-9 December 2017, PP. 3859-3869.

Shahhoseeini, R., 2009, **Evaluation of Support Vector Machines in the Classification of Hyperspectral Remote Sensing Data**, Ms.c Thesis, University of Tehran, Tehran, Iran.

Wang, Y., Sun, A., Han, J., Liu, Y. & Zhu, X., 2018, **Sentiment Analysis by Capsules**, In Proceedings of the 2018 World Wide Web Conference on World Wide Web, Lyon, France, 23-27 April 2018; PP. 1165-1174.

Xu, J.L., Esquerre, C. & Sun, D.W., 2018. **Methods for Performing Dimensionality Reduction in Hyperspectral Image Classification**, Journal of Near Infrared Spectroscopy, 26(1), PP. 61-75.

Xu, Q., Wang, D.Y. & Luo, B., 2020, **Faster Multiscale Capsule Network with Octave Convolution for Hyperspectral Image Classification**, IEEE Geosci. Remote Sens. Lett, 18(2).

Xue, Z., You, D., Candemir, S., Jaeger, S., Antani, S., Long, L.R. & Thoma, G.R., 2015, **Chest X-Ray Image View Classification**, In Proceedings of the 28th International Symposium on Computer-Based Medical Systems (CBMS) (PP. 66-71), Brazil: Ribeião Preto.

Yu, Y., Gu, T., Guan, H., Li, D. & Jin, S., 2019, **Vehicle Detection from High-Resolution Remote Sensing Imagery Using Convolutional Capsule Networks**, IEEE Geosci. Remote Sens. Lett., 16, PP. 1894-1898.

Archive of SID.ir

Archive of SID.ir

سنجش از دور
و
GIS ایران

سنجش از دور و GIS ایران سال پانزدهم، شماره سوم، پاییز 1402
Iranian Remote Sensing & GIS Vol.15, No. 3, Autumn 2023

41-60

Improving the Classification of Hyperspectral Images Using the Combined Model of CapsNet and the Extreme Gradient Boosting

Ahmadi P.[1*], Managhebi T.[2], Ebadi H.[3], Asghari Beirami B.[4]

1. Ph.D. Student, Faculty of Geomatics Engineering, K.N. Toosi University of Technology,Tehran, Iran
2. Assistant Professor, Faculty of Mapping and Spatial Information Engineering, University of Tehran, Tehran, Iran
3. Prof. Photogrammetry & Remote Sensing Dept, K.N. Toosi University of Technology, Tehran, Iran
4. Ph. D. in Remote sensing engineering, Faculty of Geomatics Engineering, K.N. Toosi University of Technology, Tehran, Iran

Abstract

With the development of remote sensing science, the use of hyperspectral images is becoming more widespread. Classification is one of the most popular topics in hyperspectral remote sensing. In the last two decades, a number of methods have been proposed to address the problem of hyperspectral data classification.In the present study, a structure based on learning capsule networks has been used to classify hyperspectral images, so that the network structure can have the most optimal generation of features by using a convolution layer and a capsule layer, and at the same time Avoid overfitting the on training data. The obtained results show the high quality of production features in the proposed structure.

In order to improve the classification accuracy, the feature extraction approach through the designed network and the classification by the Extreme Gradient Boosting was compared with the classification method by the global deep network. The proposed capsule approach consists of 3 basic layers: 1) Prime caps, which are capsules of size 8 and 32 with 9 × 9 filters and movement step 2, 2) Digitcaps with 10 16-dimensional capsules, and 3) fully connected layer. The results of examining two approaches for deep networking as well as combining capsule networks with XGBoost reinforcement tree algorithm were compared. Approaches such as SVM, RF-200, LSTM, GRU and GRU-Pretanh were considered to compare the proposed approach based on the configurations mentioned in their research.

Up in addition to the study and quality measurement of production vector deep features by the proposed method in different classifiers, the ability of deep global networks in the application of classification should also be examined. The results of examining two approaches for deep network and also combining CapsNet with XGBoost show that by using the proposed combined method, images are classified with 99% accuracy on training data and 97.5% accuracy on test data.

Keywords: Classification, Hyperspectral images, CapsNet, XGBoost, Fusion model.

* Correspondence Address: No. 1346, ValiAsr Street , Mirdamad Cross, Faculty of Geomatics Engineering, K.N. Toosi University of Technology, Tehran. Tel: 09185363604
Email: ahmadi.pouya@ut.ac.ir
https://orcid.org/0009-0001-7308-1054
https://dx.doi.org/10.48308/gisj.2023.102347

Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Archive of SID.ir