نشربه علمى بژو، شي مهندسي فناوري اطلاعات مكاني

سال هفتم، شماره دوم، تابستان ۱۳۹۸ Vol.7, No.2, Summer 2019 ۲٤۱–۲٦۳ مقاله پژوهشی

ملکک الذیوس دانشگاهت تخلاجه سر دانشگاهت برزمند بردارک

تلفیق الگوریتمهای یادگیری عمیق و فیلترهای دوطرفه با هدف استخراج ساختمان از تک تصویر نوری هوایی

مهدی خوشبرش ماسوله' ، رضا شاهحسینی'*، عبدالرضا صفری ّ

۱– دانشجوی دکتری، دانشکده مهندسی نقشهبرداری و اطلاعات مکانی، پردیس دانشکدههای فنی، دانشگاه تهران ۲– استادیار دانشکده مهندسی نقشهبرداری و اطلاعات مکانی، پردیس دانشکدههای فنی، دانشگاه تهران ۳– استاد دانشکده مهندسی نقشهبرداری و اطلاعات مکانی، پردیس دانشکدههای فنی، دانشگاه تهران

تاریخ دریافت مقاله: ۱۳۹۷/۰۲/۳۱ تاریخ پذیرش مقاله: ۱۳۹۷/۰۸/۱۲

چکیدہ

مسئله استخراج ساختمان از تکتصویر نوری هوایی با قدرت تفکیک مکانی بالا، همواره بهعنوان یکی از چالش های مهم در تهیه نقشه به شمار می آید. هدف از پژوهش حاضر، بهره گیری از قطعهبندی معنایی تک تصویر نوری هوایی بر پایه تلفیق شبکه های عصبی کانوولوشنی عمیق و فیلترهای دوطرفه جهت استخراج ساختمان می باشد. به همین منظور، پس از انتخاب یک مجموعه داده مناسب از تصاویر نوری سه باندی، با در نظر گرفتن این که برای آموزش شبکه های عصبی عمیق نیاز است تا داده های آموزشی با تعداد زیاد انتخاب شوند و باتوجه به محدودیت های سخت افزاری در این پژوهش، پس از چندین بار آزمایش، حداقل داده ای که بالاترین نرخ دقت آموزش را به دست می دهد انتخاب گردید؛ تا شبکه های کدگذار کدگشا می باشد، با استفاده از تابع بهینه سازی بر وی هوش، با بهینه سازی شبکه عصبی عمیق سیگنت که از نوع شبکه های کدگذار -کدگشا می باشد، با استفاده از تابع بهینه سازی بر وی هریش با بهینه سازی شبکه عصبی عمیق سیگنت که از نوع پردازش و استخراج ساختمان از تصاویر نوری صورت گرفته است. نتایج تحقیق بر روی مجموعه داده مربوط به تصاویر نوری هوایی از نوع بهری واقع در شهر پوتسدام آلمان از مجموعه داده های دو بعدی بر چسب دار انجمن بین المللی فتوگرامتری و سنجش از دور شبکار گیری تلفیقی شبکه عصبی عمیق بهینه سازی شده ای دو به دی برچسب دار انجمن بین المللی فتوگرامتری و سنجش از دور شاسانی ای مرز ساختمان در مقایسه با روشهای مشابه از تصاویر نوری با قدرت تفکیک مکانی بالا دارد. همچنین نتایج حاصل از روش می به کارگیری تلفیقی شبکه عصبی عمیق بهینه سازی شده سیکنت و فیلترهای دو طرفه با کرنل گوسین قابلیت های بسیار مناسبی از نظر بهبود شناسایی مرز ساختمان در مقایسه با روشهای مشابه از تصاویر نوری با قدرت تفکیک مکانی بالا دارد. همچنین نتایج حاصل از روش ساخاس این مرز ساختمان در مقایسه با روشهای مشابه از تصاویر نوری با قدرت تفکیک مکانی بالا دارد. همچنین نتایج حاصل از روش

کلیدواژهها: استخراج ساختمان، تکتصویر نوری هوایی ، قطعهبندی معنایی، شبکههای عصبی کانوولوشنی عمیق، فیلترهای دوطرفه

«نویسنده مکاتبه کننده: گروه فتوگرامتری و سنجش از دور، دانشکده مهندسی نقشهبرداری و اطلاعات مکانی، پردیس دانشکدههای فنی، دانشگاه تهران، تهران، ایران. تلفن: ۲۱۶۱۱۱۴۵۲۰

Email: rshahosseini@ut.ac.ir

سال هفتم ● شماره دوم ● تابستان ۱۳۹۸

۱– مقدمه

استخراج ساختمان از تصاویر فتوگرامتری و سنجش ازدور نقش مهمی در طیف وسیعی از کاربردهای برنامهریزی شهری و مدیریت املاک و زمین دارد [۱ و ۲]. همچنین امروزه با به کارگیری الگوریتمهای یادگیری عمیق گام مؤثری در استخراج خودکار و دقیق این عارضه مهم از تصاویر فتوگرامتری و سنجش از دور برداشته شده است، اما همواره استخراج ساختمان با ساختار هندسی پیچیده از تک تصویر نوری یک چالش محسوب می شود [۳].

پس از ارائه مدل موفق لینِت در سال ۱۹۹۳ بهعنوان اولین نمونه شبکه عصبی کانوولوشنی^۲ تا سال ۲۰۱۸ و معرفی شبکه عصبی دوئل^۳ بهعنوان یکی از ده فناوری برتر سال توسط یکی از نشریههای مؤسسه فناوری ماساچوست[†] ایالاتمتحده آمریکا، تحولات رو به رشدی در توسعه الگوریتمهای یادگیری عمیق⁶ در حیطه فناوری هوش مصنوعی⁶ رخ داده است [۴]. شبکههای عصبی عمیق دارای بیش از دو لایه مخفی^۷ هستند، بەھمین علت به آنھا شبکەھای عصبی عمیق می گویند [۵]. شبکههای عصبی عمیق اغلب مشتمل بر لایه ورودي، لايه مخفى و لايه خروجي هستند، اما لايه مخفى می تواند شامل چندلایه از ویژگیهای سطح پایین (مثل لبه ساختمان) تا ویژگیهای سطح بالا (مثل مدل هندسی ساختمان) باشد [۶]. مطالعات اخیر در حیطه یردازش تصاویر فتوگرامتری و سنجش از دور نشان میدهد، یادگیری بازنماییهای ویژگی^۸ با استفاده از شبکه عصبی کانوولوشنی عمیق در زمینههای شناسایی و استخراج عوارض دارای نتایج مطلوب و قابل اعتمادی است [۷، ۸، ۹، ۱۰، ۱۱ و ۱۲]. محورهای مهم این مطالعات اغلب شامل: اهمیت آمادهسازی صحیح دادهها در بهبود نتایج الگوریتم، استفاده از دادههای آموزشی با تعداد بالا، استفاده از یادگیری نظارتنشده جهت

پیش آموزش، طراحی یا انتخاب معماری مناسب برای شبکه عصبی عمیق با توجه به هدف اصلی مسائل، استفاده از دادههای کمکی نظیر مدل رقومی سطح حاصل از دادههای لایدار و همچنین به کارگیری شبکههای عصبی کانوولوشنی برای ادغام باندهای تصاویر چندطیفی سنجش از دور می باشد.

برای طبقهبندی تصویر، یک راهحل کارآمد جهت حل مشکل در تعمیم الگوریتمهای مختلف یادگیری ماشین استفاده از یادگیری عمیق (شبکههای عصبی كانوولوشنى عميق) است [۱۳ و ۱۴]. شبكههاي عصبي كانوولوشني عميق داراي قدرت تعميم پذيري بالايي است به نحوی که قابلیتهایی نظیر تشخیص و شناسایی اشیا [10]، استخراج ویژگی^۹ [۱۶]، تجزیهوتحلیل صداها [۱۷] و طبقهبندی تصاویر [۱۸] را با درصد موفقیت بالایی انجام میدهد. بسیاری از تحقیقات صورت گرفته در زمینه طبقهبندی و استخراج اشیا از تصاویر فتوگرامتری و سنجش از دور با استفاده از الگوریتمهای یادگیری عمیق، طی شش سال اخیر صورت گرفته است. جدول (۱) به مروری بر این مطالعات پرداخته است. با توجه به جدول (۱)، همواره مهم ترین دغدغه در مبحث به کارگیری الگوریتمهای یادگیری عمیق در پردازش تصاویر سنجشازدور با هدف استخراج یک شئ بهخصوص مانند ساختمان و یا بهینهسازی نتایج حاصل از الگوریتمهای یادگیری عمیق میباشد.

⁸ Feature Visualization

Downloaded from jgit.kntu.ac.ir on 2025-07-04

⁵ Deep Learning (DL)

⁶ Artificial Intelligence (AI)

⁷ Hidden Layer

⁹ Feature Extraction

² Convolutional Neural Networks (CNNs)

³ Dueling Neural Network (DNN)

⁴ The Massachusetts Institute of Technology (MIT)

مهدی خوش برش ماسوله و همکار ان

0.			
رهيافت تحقيق	معماری شبکه عصبی عمیق	زبان برنامەنويسى (كتابخانە)	پژوهشگر (سال)
ارائه اولین مدل موفق شبکه شش لایه عمیق جهت طبقهبندی تصاویر سنجش از دور در شش کلاس کاربری با دقتی در حدود ۷۵ تا ۹۱ درصد و بدون انجام هرگونه پیش پردازش بر روی دادههای ورودی [۱۹]	C+S+C+ S+C+FCO (TF: Tanh)	-	نگوین ٔ و همکاران (۲۰۱۳)
استفاده از شبکههای عصبی کانوولوشنی با هدف استخراج ساختمان، راه و پهنههای آبی از تصاویر سنجش از دور با قدرت تفکیک مکانی بالا با دو رویکرد پیکسل مبنا و شئ مبنا [۲۰]	C+P+C+C+ FCO +FCO (TF: ReLU+ Logistic)	C++	شو ^۲ (۲۰۱۴)
ارائه روشی نوین و مدل مبنا بهمنظور استخراج ساختمانها و شناسایی خودکار مدل سقف آنها از قبیل سقف مسطح، شیروانی، شیبدار و هرمی که در آن از شبکههای عصبی کانوولوشنی با معماری عمیق بهمنظور یادگیری کانوولوشنی با معماری عمیق استخراج شده از دادههای لایدار و تصاویر ارتوفتو استفاده شده است [11]	-	متلب (مت-کانو-نت و کفی)	علیدوست ^۳ و عارفی (۲۰۱۶)
ارائه یک شبکه عصبی کانوولوشنی به نام PNN، برای ادغام باندهای تصاویر سنجش از دور با قدرت تفکیک مکانی بالا با معماری سه لایه با تمرکز بر ادغام باندهای تصاویر FooEye-1، IKONOS و WorldView-2 [۲۲]	C+P+C+P+C+P (TF: ReLU+ ReLU+Purelin)	متلب (مت-کانو-نت و کفی)	ماسی [†] و همکاران (۲۰۱۶)
استخراج خودکار ساختمان از مدل رقومی سطح نرمال شده حاصل از زوج تصویر 2-WorldView و با استفاده از شبکه عصبی تماماً کانوولوشنی و بررسی قابلیتهای معماری پیشنهادی در مقایسه با تلفیق شبکه عصبی تماماً کانوولوشنی – میدان تصادفی شرطی و تلفیق شبکه عصبی چهار لایه – میدان تصادفی مارکفی [۲۳]	$C+C+P+C+C+P+\\C+C+C+P+C+C+\\C+P+C+C+P+\\FCN+DP+FCN+\\DP+C+Decn+SK+\\Decn+SK+Decn+\\SM+LS\\(TF: ReLU)$	پايتون (كفى)	بیتنر ^ه و همکاران (۲۰۱۷)

جدول ۱: مروری بر پیشینه تحقیق

¹ Nguyen ² Shu ³ Alidoost

⁴ Masi

⁵ Bittner

نشریه علمی پژوهشی – مهندسی فناوری اطلاعات مکانی

ارائه یک نمونه شبکه عصبی کانوولوشنی کارآمد بر اساس معماری کدگذار - کدگشا برای قطعهبندی تصویر به نام سِگنت با قابلیت قطعهبندی تصاویر مختلف و توانایی بالا در استخراج اشیا با جزئیات زیاد و پیچیده [۲۴]	$\begin{array}{c} C+BN+C+BN+P+\\ C+BN+C+BN+P+\\ C+BN+C+BN+C+\\ BN+P+C+BN+C+\\ BN+C+BN+P+C+\\ BN+C+BN+C+BN+\\ P+UP+C+BN+C+\\ BN+C+BN+UP+\\ C+BN+C+BN+C+\\ BN+UP+C+BN+C+\\ BN+C+BN+UP+\\ C+BN+C+BN+UP+\\ C+BN+C+BN+UP+\\ C+BN+C+BN+UP+\\ C+BN+C+BN+SM\\ (TF: ReLU) \end{array}$	پايتون (كفى)	بدرینایارانان ^۱ و همکاران (۲۰۱۷)
استخراج ساختمان از تصاویر هوآبرد با قدرت تفکیک مکانی بالا با ارائه معماری جدیدی بر اساس یادگیری عمیق و بهینهسازی نتایج به دست آمده با استفاده از فیلترهای هدایتی مبتنی بر ارتقاء کارایی شبکه عمیق رِز نِت با نام رِز-یو-نِت [۶]	Res-U-Net = EC+BD+DC $RN=C+BN+C+AD$ $EC=RN(1)+RN(2)+RN(3)+$ $RN(4)$ $BD=BN+C+BN+C+AD$ $DC=UP+RN(4)+UP+$ $RN(3)+UP+RN(2)+RN(1)+C$ $(TF: ReLU)$	پایتون (تنسورفلو و کراس)	ژو و همکاران ^۲ (۲۰۱۸)
قطعهبندی خودکار ساختمان در تصاویر هوایی با استفاده از شبکههای عصبی کانوولوشنی چند قیدی. این مدل بر پایه شبکه یو-نِت بنا شده است و قیود بهصورت چند مقیاسِ بین لایههای نمونه افزا به نحوی که به واقعیت زمینی مرتبط شود قرار می گیرد [۲۵] دار؛ FCO: لایه تماماً متصل؛ TF: تابع فعالساز؛ P:	C+BN+C+BN+MP+UP+ SK+C+BN+C+BN (TF: ReLU) مميق: C: لايه كانوولوشني؛ S: لايه نمونهبره	۔ ی شبکہھای عصبی ع	وو و همکاران ^۳ (۲۰۱۸) اختصارات معمار
الایه کاهش ابعاد؛ FCN: لایه تماماً کانوولوشنی؛ PD: لایه حذف تصادفی اتصالات شبکه؛ Decn: لایه معکوس کانوولوشنی؛ SK لایه پَرِش؛ SM: لایه تبدیل توزیع احتمال کلاسها؛ LS: خطای شبکه؛ UP: لایه نمونه افزا؛ BN: لایه نرمال ساز خروجیها؛ RN: شبکه باقی مانده؛ AD: لایه جمع باقی ماندهها؛ EC: شبکه کدگذار؛ BB: شبکه ارتباطی کدگذار و کدگشا؛ DC: شبکه کدگشا؛ MP			

لايه كاهش ابعاد بيشينه

برای حل مسئله استخراج ساختمان اغلب تحقیقات صرفاً و منجر می شود تا نتایج هرچند ناخواسته دچار نقص با ادغام برخی روشهای بهبوددهنده خودکار لبه (مانند شوند؛ بههمین علت نوآوریهای تحقیق حاضر در قالب فیلترهای هدایتی) در پسنتیجه الگوریتم یادگیری دو عنوان مورد بحث است که عبارت اند از: عمیق استفاده کردهاند، اما استفاده از روشهای خودکار، ۱- بهینه سازی فرآیند آموزش معماری یادگیری عمیق خود عاملی برای عدم کنترل لازم بر روی نتایج می باشد در استخراج ساختمان از تصاویر نوری هوایی که در

¹ Badrinarayanan

 $^{^{2}}$ Xu

³ Wu

بسیاری از روشها این موضوع نادیده گرفته شده است (بخش ۲-۲).

۲- بهبود نتیجه یادگیری عمیق با یک پردازش مکمل کنترلشده که در این تحقیق فیلتر دو طرفه موردنظر است. فیلتر دو طرفه با قابلیت توأمان حذف نویز و بهبود لبه میتواند نقایص آموزش یادگیری عمیق ناشی از اشکالات دستگاهی و یا دادهها را به میزان قابل توجهی بهبود دهد (بخش ٣-٣) [٢۶]. این تحقیق با هدف استخراج ساختمان از تک تصویر نوری رنگی با تلفیق الگوریتمهای یادگیری عمیق و فیلترهای دو طرفه نگارش یافته است. به همین منظور، در مرحله آمادهسازی دادهها، مجموعه داده موردنظر به دو دسته دادههای آموزشی و دادههای تست تقسیم میشوند. در مرحله بعد، شبکه عمیق سِگنت ابرای انجام فرآیند استخراج ساختمان آموزش داده می شود. لازم به ذکر است که در این مرحله برای بهینهسازی شبکه از تابع بهینهسازی برآورد لحظه تطبیقی^۲ استفاده شده است. درنهایت با استفاده از فیلتر دو طرفه با کرنل گوسین که دارای قابلیتهای بسیار مناسبی جهت بهبود لبههای اشیا است. نتایج به دست آمده از شبکه عصبی عمیق بهینهسازی میشوند و پس از آن نتایج حاصل از روش پیشنهادی با نتایج شبکه عمیق سِگنت قبل و بعد از اعمال فیلتر دو طرفه با بهرهگیری از دادههای واقعیت زمینی مورد ارزیابی قرار می گیرد.

۲- مبانی نظری

در این قسمت، ابتدا شبکههای عصبی کانوولوشنی عمیق معرفی میشوند. سپس روند بهینهسازی شبکههای عصبی کانوولوشنی عمیق بیان میشود و در نهایت فیلتر دو طرفه معرفی می گردد.

۲-۱- شبکههای عصبی کانوولوشنی عمیق با توجه به اهمیت زیاد شناخت تئوری شبکههای عصبی کانوولوشنی در این تحقیق، در این قسمت به مرور کلی

¹ SegNet

فرآیند کار این نوع از شبکههای عصبی عمیق پرداخته شده است. پایه شبکههای عصبی مصنوعی اولیه، پرسپترون است که جزء شبکههای پیشرو به شمار میآید [۲۷]. شبکههای عصبی کانوولوشنی نوع خاصی از شبکههای عصبی عمیق پیشرو^۳ هستند [۲۸]. از جمله ویژگیهای این شبکهها عبارت است از:

الف) وجود ارتباط محلی بین نورونها^۴، در شبکههای عصبی کانوولوشنی به جای متصل بودن نورونها به تمام نورونهای قبل از خود، تنها به ناحیهای کوچک از داده ورودی (تصویر چند باندی) متصل است. نکته مهم در این شبکهها این است که اتصالات علاوه بر اینکه طول و عرض (سطح تصویر) را پردازش میکنند، بعد سوم یعنی باندهای تصویر را نیز مورد پردازش قرار میدهند، به همین دلیل این شبکهها برای پردازش تصاویر چند طیفی بسیار موردتوجه هستند؛

ب) وجود وزنهای مشترک در هر لایه، جهت کنترل
 تعداد پارامترهای مورد استفاده در لایههای کانوولوشنی
 از وزنهای مشترک در هر لایه استفاده می شود. در عمل
 بهره گیری از چنین ویژگی، موجب کاهش تعداد
 پارامترها در شبکه می شود؛

پ) بهرهگیری از تعداد زیادی لایه مخفی؛ که موجب افزایش تولید ویژگیهای جدید از پایینترین سطح ویژگی مثل لبهها تا بالاترین سطح ویژگی مثل مدلهای هندسی میشود.

ت) به کارگیری لایه کاهش ابعاد^۵، استفاده از لایه کاهش ابعاد در بین لایههای کانوولوشنی می تواند موجب کاهش اندازه مکانی (طول و عرض) دادههای ورودی و در نهایت منجر به کاهش پارامترهای شبکه شود. استفاده از این لایهها قابلیت کنترل بر بیش برازش دادهها را به شبکه می دهد [۲۹]. در شبکههای عصبی کانوولوشنی عمیق مجموع دادههای ورودی وزندار در هر لایه از تابعی غیرخطی عبور می کند. خروجی این مرحله همان

² Adaptive Moment Estimation (ADAM)

³ Deep Feedforward Networks

⁴ Local Connectivity

⁵ Pooling

سال هفتم ● شماره دوم ● تابستان ۱۳۹۸

ویژگیهای استخراجشده از دادههای ورودی میباشد. در این شبکهها مجموعهای از لایههای کانوولوشنی در کنار تعداد لایه چگال با اتصال کامل، کار استخراج ویژگی و طبقهبندی را انجام میدهند و فرآیند بهروزرسانی وزنها نیز همانند دیگر شبکههای عصبی مبتنی بر قاعده بازگشتی میباشد. در شبکههای عصبی کانوولوشنی مسئله کاهش سرعت آموزش شبکه از جمله موارد مهم بشمار میرود، که دلیل آن هم تغییر پارامترها در طول عبور از لایههای مختلف میباشد. از جمله راهکارهای مناسب جهت حل مشکل کند بودن سرعت آموزش ماستفاده از روش حذف تصادفی توصیه شده است. که در این روش نتیجه برخی از نورونها به صورت تصادفی از فعالیت آموزش حذف میشود که موجب عدم وابستگی زیاد مدل به دادههای آموزشی میشود.

۲-۲- بهینهسازی شبکه عصبی کانوولوشنی عمیق بااستفاده از الگوریتم بر آورد لحظه تطبیقی

در شبکههای عصبی بهینهسازی پارامترهای وزن و بایاس شبكه از جمله موضوعات بسيار مورد اهميت است. الگوریتمهای یادگیری عمیق اغلب با بهکارگیری روشهای مختلف نظیر گرادیان نزولی، بهینهسازی می شود. هدف روشهای بهینه سازی، بهبود عملکرد آموزش یادگیری الگوریتم (همگرایی) بهواسطه آزمون و خطای شبکه در طی فرآیند یادگیری است. همچنین جهت بهروزرسانی پارامترهای شبکه نظیر وزن و بایاس نیز مورد استفاده قرار می گیرند. لازم به ذکر است که فرآیند انجام بهینهسازی در هر بار دوره تکرار مشتمل بر: ۱) محاسبه خطای مدل طراحی شده؛ ۲) محاسبه گرادیانها (g_t)؛ ۳) اعمال روش بهینهسازی جهت بهروزرسانی پارامترهای مدل با استفاده از گرادیانها؛ ۴) انتقال مقادیر اولیه وزنها با توجه به نرخ یادگیری و الكوريتم مورد استفاده جهت بررسي وضعيت بهينهسازي مدل می باشد [۳۰]. الگوریتم برآورد لحظه تطبیقی، روشی بر مبنای گرادیان جهت بهینهسازی پارامترهای مدل است [۳۱]. اگر مقدار حرکت را همانند یک توپی که از یک سطح شیبدار به

یایین می آید در نظر بگیریم، روش بهینه ساز بر آور د لحظه تطبیقی همچون یک توپ سنگین است که با سطح اصطکاک دارد، به عبارت دیگر الگوریتمهای بهینهساز با هدف كمينه كردن تابع زيان، كار مي كنند. روش برآورد لحظه تطبیقی در سطح خطا، مینیممهای صاف (در مقابل مینیممهای تیزتر) را در نظر می گیرد [۳۲]. برمبنای محاسبات انجامشده در این تحقیق به کار گیری بهینهساز برآورد لحظه تطبیقی نرخ همگرایی را بهبود مىدهد. پارامترهاى بهينەساز برآورد لحظه تطبيقى مشتمل بر η نرخ یادگیری، β_1 ترم تأخیری میانگین heta گرادیان، β_2 ترم تأخیری میانگین مربعات گرادیان، θ بردار پارامترها، $f(\theta)$ تابع هدف تصادفی بر مبنای θ و پارامتر ϵ که جهت جلوگیری از میل به صفر کردن مقدار بردار پارامترها در هر مرحله بهروزرسانی تعریف میشود. قانون بهروزرسانی بردار پارامترها (وزن و بایاس شبکه) بهصورت رابطه (۱) تعريف مي شود.

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \varepsilon} \, \hat{m}_t \tag{1}$$

در رابطه (۱) مقدار \hat{m}_t برآورد مقدار حرکت در لحظه اول (میانگین) و مقدار \hat{v}_t برآورد مقدار حرکت در لحظه دوم (توزیع غیر مرکزی واریانس) میباشد، که بهصورت رابطه (۲)و (۳) محاسبه می شوند.

$$\hat{m_t} = \frac{m_t}{1 - \beta_1^t} \tag{(Y)}$$

$$\hat{v_t} = \frac{v_t}{1 - \beta_2^t} \tag{(7)}$$

در عمل مقادیر \hat{m}_t و \hat{v}_t نقش تصحیح بایاس ها در بردار پارامترها را دارند. مقادیر m_t و v_t به ترتیب برابر میانگین گرادیان و میانگین مربعات گرادیان میباشند. که با استفاده از رابطه (۴) و (۵) تعریف می شوند.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$
 (4)

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) g_t^2$$
 (۵)

مقادیر پیشنهادی در یک پژوهش برای $\beta_2 \beta_1 \beta_2 \beta_1$ برابر با γ_1 ، γ_2 مقادیر γ_1 ، γ_2 مقادیر می می مناسبی می اشند [۳۲].

مهدی خوش برش ماسوله و همکار ان

لحظه تطبیقی استفاده شده است. فرآیند پیادهسازی این

الگوريتم با زبان متلب توسط تابع زير صورت گرفته است.

در این پژوهش، جهت بهینهسازی پارامترها و فرآیند آموزش و تست مدل عمیق سِگنت از الگوریتم برآورد

الگوريتم برآورد لحظه تطبيقي

Function [w, state] = adam (teta, state, grad, opts, lr)

- 1. if nargin == 0 % Returns the default solver options
- 2. w = struct ('beta1', 0.9, 'beta2', 0.999, 'eps', 1e-8);
- 3. return;
- 4. end
- 5. *if isequal(state, 0) % start off with state = 0 so as to get default state*
- 6. state = struct('m', 0, 'v', 0, 't', 0);
- 7. end
- 8. state.m = opts.beta1 * state.m + (1 opts.beta1) * grad; % update first moment vector `m`
- 9. state.v = opts.beta2 * state.v + (1 opts.beta2) * grad. ^2; % update second moment vector `v`
- 10. state.t = state.t + 1; % update the time step
- 11. lr_t = lr * (((1 opts.beta2^state.t) ^0.5) / (1 opts.beta1^state.t)); % biased estimates of first and second moment vectors
- 12. $teta_t+1 = teta_t lr_t * state.m. / (state.v. ^0.5 + opts.eps); % Update teta$

وزنها فقط بر اساس فاصلهی اقلیدسی تعیین نمی شوند، بلکه تفاوتهای رادیومتریکی نیز در آن تأثیر گذار است. این فیلتر با تنظیم کردن وزنهای پیکسلهای همسایه در یک حلقه بهصورت سامانمند، لبههای تیز تصاویر را حفظ می کند [۳۳ و ۳۴]. فیلتر دو طرفه بهصورت رابطه(۶) و رابطه تبدیل نرمال بهصورت رابطه (۷) تعریف می شود.

$$\operatorname{Im}^{filtered} = \frac{1}{W_P} \sum_{x_i \in \Omega} \operatorname{Im}(x_i) f_r(|\operatorname{Im}(x_i) - \operatorname{Im}(x)|) g_s(|x_i - x|)$$

$$W_{p} = \sum_{x_{i} \in \Omega} f_{r} \left(\left| \operatorname{Im}(x_{i}) - \operatorname{Im}(x) \right| \right) g_{s} \left(\left| x_{i} - x \right| \right)$$

در این تحقیق از فیلتر دو طرفه برای بهینه سازی خروجی پردازش با الگوریتم یادگیری عمیق استفاده شده است. با توجه به شکل (۱) با فرض ثابت بودن ابعاد کرنل (در این تحقیق از ابعاد ۹×۹ استفاده شده است)، با اعمال مقادیر مختلف بر روی نتیجه به دست آمده از پردازش الگوریتم یادگیری عمیق پیشنهادی تحقیق، مقدار بهینه برای تابع گوسین در حیطه مکان (g_s) برابر ۸ و مقدار بهینه برای تابع گوسین در حیطه دامنه (f_r) برابر ۵۰٬۰ ۲-۳- فیلترهای دو طرفه فیلتر دو طرفه^۱ یک فیلتر غیرخطی ملایم ساز با حفظ لبه و کاهش نویز برای تصاویر است. در این فیلتر، میزان شدت هر پیکسل با میانگین وزندار پیکسلهای همسایه جایگزین میشود که این وزن میتواند بر اساس توزیع نرمال گوسین باشد. نکته مهم در این فیلتر آن است که

رابطه(۷)

¹ Bilateral Filter (BF)

سال هفتم ● شماره دوم ● تابستان ۱۳۹۸

موردنظر بهعنوان ورودی، محاسبه فاصله گوسی بر اساس موقعیت مرکز فیلتر بر روی تصویر (X, Y)، استخراج ناحیه محلی بر مبنای تبدیل نرمال و در نهایت فیلتر کردن تصویر بر اساس مقادیر محاسبه شده با تلفیق در حیطه مکان و دامنه می باشد. بهترین نتایج را برای بهبود لبههای ساختمانهای استخراجشده ارائه میدهند. لازم به ذکر است که افزایش ابعاد کرنل موجب پدید آمدن آثار مخرب روی تصویر میشود و همچنین اگر ابعاد کرنل کوچک انتخاب شود نتیجه مطلوب نخواهد بود [۳۶].

فرآیند پیادهسازی فیلتر دو طرفه بر اساس شبه کد زیر، مشتمل بر تعریف پارامترهای فیلتر دو طرفه و تصویر



شکل ۱: اعمال فیلتر دو طرفه روی تصویر (p و p متناظر با همان مختصات مرکزی فیلتر x در رابطه (۶) میباشند) [۳۶]

الگوريتم فيلتر دو طرفه

Bilateral_Filter (Im, w_p, g_s, f_r) % Bilateral filter Parameters

- 1. Obtain [X,Y] = meshgrid(-w_p: w_p, w_p: w_p) % Pre-compute Gaussian Distance Weights
- 2. $G = exp(-(X.^2+Y.^2)/(2*g_s^2))$ %Compute the Gaussian Distance Weights
- 3. dim = size(Im) % Apply Bilateral Filter
- 4. B = zeros(dim)
- 5. Repeat for i = 1:dim (1)
- 6. *Repeat for* j = 1: *dim* (2)
- 7. *iMin* = max (*i* w_p,1) % Extract Local Region
- 8. $iMax = min(i + w_p, dim(1))$
- 9. $jMin = max(j w_p, 1)$
- 10. $jMax = min(j + w_p, dim(2))$
- 11. I = Im (iMin:iMax,jMin:jMax)
- 12. $H = exp(-(I Im(i,j)).^2/(2 * f_r^2))$
- 13. F = H.*G((iMin:iMax)-i+ w_p +1,(jMin:jMax)-j+ w_p +1) % Calculate Bilateral Filter Response
- 14. B(i,j) = sum(F(:), *I(:))/sum(F(:))
- 15. end
- 16. end

۳- روش تحقيق

با توجه به شکل(۲)، این تحقیق در چهار مرحله پیادهسازی و مورد ارزیابی قرار گرفته است. مرحله اول، شامل تهیه مجموعه تصاویر نوری رنگی هوابرد با قدرت تفکیک مکانی بالا است. مرحله دوم، شامل آمادهسازی، تقسیم و استخراج نمونههای آموزشی و تست برای اجرای بهتر مدل موردنظر میباشد. مرحله سوم، شامل طراحی و بهینهسازی شبکه عصبی عمیق سگنت است که با استفاده از نمونههای آموزشی مرحله قبل، آموزش میبیند و پس از انجام مرحله اعتبارسنجی، با استفاده از دادههای تست ارزیابی میشود. در مرحله چهارم نتیجه بهدست آمده از روش پیشنهادی با نتایج حاصل از روشهای *Res-U-Net+GF* [۶]، الگوریتم پایه سگنت *CNN+RF+CRF* [۶]، الگوریتم پایه سگنت *CNN+RF+CRF* [۳] و ۲۸۲+*CN* (۳7] با استفاده از نقشههای واقعیت زمینی و برمبنای معیارهای ارزیابی دقت مورد بررسی قرار میگیرند.

تلفیق الگوریتمهای یادگیری عمیق و فیلترهای ...

مهدی خوش برش ماسوله و همکار ان

۳–۱– مجموعه داده تحقيق

شکل(۳) مجموعه داده مورد استفاده در این تحقیق مربوط به تصاویر نوری رنگی هوابّرد با قدرت تفکیک مکانی بالا از مناطق شهری واقع در شهر پوتسدام آلمان از مجموعه دادههای دو بعدی برچسبدار انجمن بینالمللی فتوگرامتری و سنجش از دور¹ را نشان میدهد [۳۸]. تصاویر مورد استفاده شامل سه باند قرمز، سبز و آبی با قدرت تفکیک مکانی ۵ سانتیمتر میباشد. این مجموعه داده شامل ۳۸ تصویر قائم که هرکدام از این محموعه داده شامل ۳۸ تصویر قائم که هرکدام از این تصاویر دارای ابعاد ۶۰۰۰×۶۰۰۰ پیکسل به همراه دادههای واقعیت زمینی برچسب دار میباشد. از این دادههای در چندین تحقیق اخیر در حیطه یادگیری عمیق برای آموزش و تست الگوریتمهای مختلف استفاده شده است [۶، ۳۹، ۴۰ و ۴۱].



شكل ٢: الگوريتم پيشنهادى تحقيق

¹ International Society for Photogrammetry and

سال هفتم • شماره دوم • تابستان ۱۳۹۸



شکل ۳: منطقه مورد مطالعه (شکل بالا از راست به چپ: موقعیت کشور آلمان (سبز پررنگ) در میان کشورهای اتحادیه اروپا (سبز کمرنگ) و سایر کشورهای همجوار (خاکستری) [۴۲ و ۴۳]، موقعیت شهر پوتسدام در میان شهرهای مجاور؛ شکل پایین: ترتیب بلوکهای عکسبرداری از شهر پوتسدام)

۲-۳- آمادهسازی دادهها

با توجه به حجم بالای تصاویر و عدم امکان پردازش تمامی این مجموعه داده به علت محدودیتهای سختافزاری، در این تحقیق از هشت بلوک (جدول (۲)) برای آموزش شبکه که ده درصد از این دادهها برای اعتبارسنجی مدل مورد استفاده قرار می گیرد و از سه بلوک برای تست شبکه استفاده شده است. این بلوکها با در نظر گرفتن اینکه در بر گیرنده مناطق ساختمانی

بیشتر و متنوعتری (از نظر هندسه و نوع سقف) باشد در نظر گرفته شده است. همچنین جهت بهبود عملکرد شبکه، هر یک از بلوکها به تصاویری با ابعاد ۱۲۸×۱۲۸ تقسیم میشوند. علاوهبرآن، مناطق ساختمان و غیرساختمان در واقعیت زمینی بهوسیله ماسککردن ساختمانها از سایر کلاسها جدا شدند و سپس مورد استفاده قرار گرفتند.

جدول ۲: تعداد و شماره بلوکهای مناطق آموزشی و تست

تعداد	شماره بلوکها (رديف)	
۸ بلوک	۱۰ و ۱۲ (۳) – ۱۰ و ۱۱ (۴) ۱۱ و ۱۲ (۵) – ۸۰ و ۰۹ (۶)	مناطق آموزشي
۳ بلوک	(Y))) - (۶) • λ - (۲)) •	مناطق تست

به دلیل هزینه محاسباتی بالا در به کارگیری یک تصویر با ابعاد بزرگ به صورت یکجا در فرآیند آموزش، موضوع تقسیم بندی داده آموزشی و ارائه یک برنامه تغذیه از داده آموزشى براى آموزش الگوريتم يادگيرى عميق بسيار مورد اهمیت است. به همین منظور با تستهای جزئی توان یردازش سیستم مورد استفاده آزمایش گردید و در نهایت تقسیم بندی دادهها با ابعاد سطحی ۱۲۸×۱۲۸ فرآیند را تسریع بخشید و روند پیادهسازی را سهلتر ساخت. شکل(۴) سلسلهمراتب استخراج ویژگی در لایههای کانوولوشن و لایه کاهش ابعاد بر روی یک قطعه تصویر رنگی با ابعاد ۱۲۸×۱۲۸×۳ را نشان میدهد. شکل(۴) روند پردازش بر روی تصویر ورودی با استفاده از لایه کانوولوشنی و لایه کاهش ابعاد را نشان میدهد. در لایه کانوولوشنی با جاروب کردن تصویر در هر باند توسط فیلتر با ابعاد مشخص (با توجه به جدول (۳) در این تحقیق ابعاد فیلترها ۳×۳ است) و سپس استفاده از لايه كاهش ابعاد تصوير كاهش مييابد (همانطور كه مشخص است همه لايهها به يكديگر متصل نيستند) و در یک بردار ویژگی جدید جمع می شوند. این روند با توجه به عُمق شبکه (تعداد لایههای مخفی) ادامه مىيابد.

۳-۳- پردازش با الگوریتمهای یادگیری عمیق

در این تحقیق برای انجام فرآیند استخراج ساختمان از تصاویر هوابُرد با قدرت تفکیک مکانی بالا از معماری سِگنت استفاده شده است [۲۴]. علت استفاده از این معماری، قابلیت بالای استخراج لبهها و همچنین تشخیص اشیا با جزئیات پیچیده و سازگاری با تصاویر رنگی بهعنوان ورودی شبکه است [۴۴ و ۴۵]. این معماری عمیق در محیط نرمافزار متلب ۲۰۱۷ و با استفاده از جعبهابزار مت-کانو-نت طراحی و بر روی مجموعه داده تحقیق اجرا شده است. کلیه فرآیند پردازش اعم از آموزش و تست مدل و اعمال فیلتر دو طرفه، در این تحقیق مبتنی بر پردازش بر روی کارت

[\] Batch Normalisation (BN)

تلفیق الگوریتمهای یادگیری عمیق و فیلترهای ... مهدی خوش برش ماسوله و همکار ان

گرافیک GeForce GTX 550 Ti صورت گرفته است. شکل(۵) معماری شبکه سِگنت و فرآیند پردازش و قطعهبندی تصویر را نشان میدهد. در این معماری با حذف لایه تماماً متصل در سراسر شبکه و استفاده از لایههای کانوولوشنی (آبی)، و بهرهگیری از لایههای کدگذار و کدگشا فرآیند قطعهبندی تصویر انجام میشود. در این شبکه جهت بهبود نتایج مرحله کدگذاری از لایه نرمال ساز خروجیها^۱ استفاده شده است، همچنین تابع فعال ساز در این شبکه تابع رلو میباشد. بهرهگیری از لایههای نمونه افزا در مقابل لایههای کاهش ابعاد، نیز ایده دیگر این معماری میباشد.



شکل ۴: فرآیند سلسله مراتبی استخراج ویژگی با استفاده از شبکههای عصبی کانوولوشنی عمیق



شکل ۵: معماری شبکه سِگنت (نتیجه داده تست سه، بدون اعمال فیلتر دو طرفه)

سال هفتم ● شماره دوم ● تابستان ۱۳۹۸

نشریه علمی پژوهشی – مهندسی فناوری اطلاعات مکانی

در هر یک از قسمتهای این شبکه فرآیند کار به این صورت است که بردارهای ویژگی استخراجشده از لایهی قبل، به پنجرههای مجزا از هم تقسیم میشوند و بیشترین مقدار موجود در این پنجره محاسبه شده و بهعنوان مقدار در بردار ویژگی جدید انتخاب میشود. لایههای نمونه افزایی شبکههای تماماً کانوولوشنی یادگیری را با معکوس کانوولوشن از بردار ویژگی ورودی انجام میدهند و سپس به بردار ویژگی کدگذاری شده اضافه میشوند و خروجی لایه کدگشا را تولید میکنند. این بردار خروجی از شاخص کاهش ابعاد بیشینه^۱ (که شامل لایه نمونه افزایی نیز میشود) در لایه کدگذار متناظر است.

در جدول(۳) جزئیات مربوط به پارامترهای هریک از لایهها در این شبکه عمیق، نشان داده شده است. برای رسیدن به دقت بالا و همگرا شدن شبکه، لازم است



تا شبکه در تعداد تکرار زیاد آموزش داده شود و پارامترهای آن با استفاده از بهینهساز برآورد لحظه تطبیقی تنظیم گردد. شکل(۶) تغییرات دقت را بر روی مجموعه داده تحقیق با افزایش تعداد تکرار نشان می دهد. بهترین مدل در تکرار ۲۹ به دست آمده است. پس از استخراج ساختمان با استفاده از شبکه عصبی پس از استخراج ساختمان با استفاده از شبکه عصبی عمیق سگنت، لازم است تا فیلتر دو طرفه به نتیجه اعمال شود تا خروجی موردنظر تحقیق به دست آید، که به آن شود تا خروجی موردنظر تحقیق به دست آید، که به آن *شود* تا خروجی می گوییم. مجموعه نتایج به دست آمده بر روی سه تصویر تست با نتایج روشهای -UNX+RF Res-U- الگوریتم پایه SegNet در SegNet به روی سه تصویر تست با نتایج روشهای -UNX+RF روی سه تصویر تست با نتایج روشهای -UNX+RF روی سه تصویر تست با نتایج روشهای -ux-RF



شکل ۶: منحنی تکرار بر روی مجموعه داده تحقیق با استفاده از روش پیشنهادی، محور افقی: تعداد تکرار، محور عمودی: اختلاف اندازهگیریها. شکل راست: خطای آموزش و خطای اعتبارسنجی در ۷۹ تکرار (بهترین مدل برای مجموعه داده تحقیق)، شکل چپ: منحنی دقت آموزش مدل

[\] Max-pooling

	ى سبعه سِمع پيسم	بحدول ۱۰ بحرميات ششار	
داده ورودی	ابعاد کرنل	گام	بردار ویژگی
Conv1_1+BN1_1+ReLU1_1	۳×۳	١	84
Conv1_2+BN1_2+ReLU1_2	۳×۳	١	84
Pool1 (Max_Pooling)	۲×۲	٢	
Conv2_1+BN2_1+ReLU2_1	۳×۳	١	١٢٨
Conv2_2+BN2_2+ReLU2_2	۳×۳	١	١٢٨
Pool2 (Max_Pooling)	۲×۲	٢	
Conv3_1+BN3_1+ReLU3_1	۳×۳	١	۲۵۶
Conv3_2+BN3_2+ReLU3_2	۳×۳	١	۲۵۶
Conv3_3+BN3_3+ReLU3_3	۳×۳	١	208
Pool3 (Max_Pooling)	۲×۲	٢	
Conv4_1+BN4_1+ReLU4_1	۳×۳	١	۵۱۲
Conv4_2+BN4_2+ReLU4_2	۳×۳	١	۵۱۲
<i>Conv4_3+BN4_3+ReLU4_3</i>	۳×۳	١	۵۱۲
Pool4 (Max_Pooling)	۲×۲	٢	
Conv5_1+BN5_1+ReLU5_1	۳×۳	١	1.74
Conv5_2+BN5_2+ReLU5_2	۳×۳	١	1.74
Conv5_3+BN5_3+ReLU5_3	۳×۳	١	1.74
Pool5 (Max_Pooling)	۲×۲	٢	
Upsample1	۲×۲	٢	
De(Conv5_3+BN5_3+ReLU5_3)	۳×۳	١	1.74
De(Conv5_2+BN5_2+ReLU5_2)	۳×۳	١	1.74
De(Conv5_1+BN5_1+ReLU5_1)	۳×۳	١	۵۱۲
Upsample2	۲×۲	٢	
De(Conv4_3+BN4_3+ReLU4_3)	۳×۳	١	۵۱۲
De(Conv4_2+BN4_2+ReLU4_2)	۳×۳	١	۵۱۲
De(Conv4_1+BN4_1+ReLU4_1)	۳×۳	١	۲۵۶
Upsample3	۲×۲	٢	
De(Conv3_3+BN3_3+ReLU3_3)	۳×۳	١	۲۵۶
De(Conv3_2+BN3_2+ReLU3_2)	۳×۳	١	۲۵۶
De(Conv3_1+BN3_1+ReLU3_1)	۳×۳	١	۱۲۸
Upsample4	۲×۲	٢	
De(Conv2_2+BN2_2+ReLU2_2)	۳×۳	١	١٢٨
De(Conv2_1+BN2_1+ReLU2_1)	۳×۳	١	۶۴
Upsample5	۲×۲	٢	
De(Conv1_2+BN1_2+ReLU1_2)	۳×۳	١	54
De(Conv1_1+BN1_1+ReLU1_1)	۳×۳	١	٢
Softmax+Loss			

جدول۳: جزئیات معماری شبکه سِگنت پیشنهادی در این تحقیق

نشریہ علمی پژوهشی – مهندسی فناوری اطلاعات مکانی

مهدی خوش برش ماسوله و همکار ان



شکل ۷: نتایج بصری تحقیق و مقایسه آن با سایر روشها

سال هفتم ● شماره دوم ● تابستان ۱۳۹۸

۳-۴- ارزیابی عددی نتایج

پس از استخراج ساختمانها از دادههای تست موردنظر با استفاده از روش پیشنهادی تحقیق، در مرحله لازم است تا نتیجه به دست آمده با دادههای مرجع (واقعیت زمینی) مقایسه شوند تا دقت نهایی خروجیهای به دست آمده مورد تایید قرار گیرد. فرآیند ارزیابی دقّت و صحّت نتایج تحقیق با دادههای مرجع با ایجاد ماتریس خطا برای تعداد اشیا ساختمانی استخراجشده صحیح و ناصحیح بررسی می شود. کلاس های ماتریس خطا برای

این ارزیابی شامل دو کلاس ساختمان و غیرساختمان میباشد. مقادیر هر درایه این ماتریس مبتنی بر تعداد پیکسلهای برچسب خورده می باشد. جدول (۴) ساختار کلی ماتریس خطا را نشان میدهد. در این جدول، TP تعداد پیکسلهای ساختمانی که ساختمان، FP تعداد پیکسلهای ساختمانی که غیرساختمان، FN تعداد پیکسلهای غیرساختمانی که ساختمان، TN تعداد پیکسلهای غیرساختمانی که غیرساختمان تشخیص داده شدهاند، میباشند.

جدول۴: ماتریس خطا

	ساختمان	غيرساختمان
ساختمان	ТР	FP
غيرساختمان	FN	TN

بر اساس ماتریس خطا، معیارهای تمامیّت^۱، صحّت^۲، و IoU و $F_{measure}$ استخراجشده بهصورت رابطه(۸) تا رابطه(۱۱) تعریف می شوند [۷ و ۲۳].

در رابطه (۱۰) جهت سادهتر کردن محاسبات مقدار β را برابر با یک در نظر می گیریم [۲۸]. در رابطه (۱۱) مقادیر

npred و ngt به ترتیب مقادیر پیکسلهای پیشبینی شده با شبکه عصبی پیشنهادی و واقعیت زمینی میباشند. نتایج این ارزیابی عددی در جدول(۵) ارائه شده است. در جدول(۵)، بهترین مقادیر به صورت برجسته نشان داده شده است.

TP

رابطه(۸)

(ابطه(٩)

 $Completeness = \frac{1}{TP + FN}$ TP

$$Correctness = \frac{TT}{TP + FP}$$

$$F_{measure} = \frac{(1+\beta^2)TP}{(1+\beta^2)TP + \beta^2 FN + FP}$$
(۱・)

$$IoU = \frac{TP}{n_{pred} + n_{gt}}$$
 (۱۱) رابطه (۱۱)

² Correctness

¹ Completeness

مهدی خوش برش ماسوله و همکار ان

		تمامیت	صحت	Fmeasure	IoU
	SegNet	٩٠٫١١	۸۹٬۸۴	٨٩٫٨۶	۵۱,۱۳
	FCN	٨٩٫٩٣	٨٩٫٧٢	٨٩٫۶٠	۵۰٫۷۷
داده تست <i>آ GF</i>	Res-U-Net+GF	१٣/٩٧	٩١,١٣	۹ ۱٬۲۸	$\Delta \Delta_{i} \wedge 1$
[(٢) ١٠]	CNN+RF	٩ ١, ۴ ٠	۸۹٬۸۵	٨٩٫۶٧	۵۲٬۳۵
RF	CNN+RF+CRF	٩ ١,۴٨	٨٩٬٩٨	٨٩٫٨٠	57,47
F	SegNet + BF	٩۵/۱۴	٩٢,٣٧	٩١/۴٩	54/84
	SegNet	٨٩٫۶٧	۸٧,١۵	٨٦,٤۵	48,71
	FCN	۸۹٫۴۰	<i>እዮ_/</i> ۶۹	٨٦,٠٢	48/11
داده تست <i>F</i>	Res-U-Net+GF	٩ ١,٢٢	٨٩٫٨۴	٨٩٫٢٨	49,89
[(۶) ·λ]	CNN+RF	٨٩٫٠۴	٨٦/٢٧	٨۵,۴۱	47/17
RF	CNN+RF+CRF	٨٩٫٢٢	۸۶ _/ ۵۶	٨۶٫۱۹	۴۵,۵۷
F	SegNet + BF	٩١,۶٧	٩+,+۲	19,44	49,V+
	SegNet	٩٣٫١۴	۹۱,۰۴	٩٠,٧۵	۵۳/۱۴
	FCN	٩ <i>٢</i> ,٩٠	۹٠,۵٠	٩٠٫٢١	۵۲,۵۵
داده تست ۳	Res-U-Net+GF	۹۵ _/ ۷۱	٩٣٫٣٣	۹۱٫۶۷	۵۸,۰۰
[(Y) 11]	CNN+RF	٩٢,٩۵	٩٠,٣۴	٩٠,٢٩	۵۲/۱۸
RF	CNN+RF+CRF	१٣/۴٩	٩٠,۶۵	٩٠,۴١	۵۲٬۶۶
F	SegNet + BF	98,14	٩٣٬٩٨	٩٢,۵٧	۵۸,۳۴

جدول ۵: نتایج استخراج ساختمان از سه منطقه تست با استفاده از روش پیشنهادی تحقیق در مقایسه با سایر روشها (برحسب ٪)

۳-۵- آنالیز کیفیت نتایج

در این قسمت جهت بررسی قابلیتهای الگوریتم پژوهش در مقایسه با الگوریتمهای مشابه در حل مسئله استخراج ساختمان از دادههای تست تحقیق دو محدوده سبز و قرمز برای آنالیز کیفیت در هر تصویر مطابق شکل (۸) انتخاب شد. با توجه به شکل (۸)، محدوده سبز رنگ برای به چالش کشیدن الگوریتم در مورد، وجود فاصلههای بسیار کوچک بین ساختمانها و تشخیص یا عدم تشخیص این فواصل توسط الگوریتم پیشنهادی انتخاب شد. در داده تست یک، اگرچه مرز بسیار باریک بین دو ساختمان با استفاده از روش پیشنهادی تحقیق از بین رفته است اما هندسه ساختمان نسبت به سایر روشها دارای نتیجه بهتری در شناسایی مرز داشته اما

هندسه ساختمان بهخوبی قابل تشخیص نیست و در مورد سایر روشها نیز خروجیها بهمراتب ضعیف هستند. در داده تست دو نیز این مسئله وجود دارد، اما الگوریتم پیشنهادی تنها مرز بین دو ساختمان کوچک را نتوانسته است تشخیص دهد که نسبت به نتیجه داده تست یک شرایط بهتر است. همچنین با توجه به نتایج سایر روشها می توان گفت استفاده از آنها، نتیجه ضعیف تری داشته است. در داده تست سه، الگوریتم تا حد زیادی این مرز کوچک را تشخیص داده و پس از اعمال فیلتر دو طرفه نیز همچنان این مرز مشخص بوده و نتیجه مناسب تری ارائه شده است و دلیل آنهم می تواند وجود بیشتر ساختمانهایی با سقف قرمز در دادههای آموزشی باشد.



شکل ۸: انتخاب محدودههای آزمایشی برای آنالیز کیفیت نتایج الگوریتم پیشنهادی

سال هفتم • شماره دوم • تابستان ۱۳۹۸

در مجموع با توجه به نتایج به دست آمده پیرامون تشخیص فواصل کوچک بین ساختمانها که مسئله مهمی در تهیه نقشه میباشد، الگوریتم پیشنهادی تحقیق در این امر موفق عمل میکند اما به دلیل وجود پوششهای گیاهی نظیر درختان کوچک و چمنها در میان این ساختمان موجب کاهش دقت این الگوریتم شده است. بهعلاوه روشهای مشابه به دلیل ساختاری که دارند اغلب نیازمند منابع داده بیشتر هستند که همین عامل باعث کاهش دقت آنها میشود.

محدوده قرمز رنگ در شکل (۸)، جهت بررسی قابلیت الگوريتم در تشخيص ساختمانهاي بسيار كوچك و همچنین تشخیص لبه ساختمانها در محدودههای پرتراکم میباشد. در داده تست یک، تعدادی از ساختمانها با وسعت بسيار كم تشخيص داده نشدهاند، اما نه همه آنها بلکه اکثر آنهایی که در کنار ساختمانهای بلند هستند و یا پوششهای گیاهی آنها را احاطه کردهاند. اما در همین محدوده شاهد عملکرد مناسب الگوريتم پيشنهادي پژوهش پس از اعمال فيلتر دو طرفه در محدوده لبه ساختمانها هستیم. در میان سایر روشهای مشابه نیز روش Res-U-Net+GF دارای نتایج نسبتاً مناسبی بوده است. در داده تست دو، ساختمان های کوچک تشخیص داده شدهاند اما همه جزئیات آن مشخص نیست. این نمونه از جمله اندک ساختمانهایی است که به دلیل پیچیدگی هندسی ساختمان و وسعت كم أن بهخوبي تشخيص داده نشده است، همچنین مشخص است که لبه ساختمانها بهبود زیادی پیدا کرده است. در نهایت در داده تست سه، مسئله عدم تشخيص ساختمانهای کوچک به حداقل میرسد و تنها بعد از اعمال فیلتر یک پیکسل در محدوده قرمز رنگ غیرقابل تشخیص می شود و لبه ساختمان دارای بهبود می شود که در مقایسه با سایر الگوریتمها، روش CNN+RF+CRF تا حدودی توانسته است به نتایج روش پیشنهادی نزدیک شود و عملکرد مناسبی داشته است.

تلفیق الگوریتمهای یادگیری عمیق و فیلترهای ...

مهدی خوش برش ماسوله و همکار ان

۴- نتیجهگیری

مسئله استخراج ساختمان از تکتصویر نوری رنگی با قدرت تفکیک مکانی بالا، همواره بهعنوان یکی از چالشهای مهم در تهیه نقشه به شمار میآید. این موضوع از این جهت دارای اهمیت است که دسترسی بسیاری از کشورها به دادههای نوری بهمراتب بیشتر و آسان تر است و می توانند فرآیند تهیه نقشه را با منبع دادهای نسبتاً ارزان و با پوشش کامل از مناطق مختلف در اختیار داشته باشند. علاوه بر آن به کار گیری دادههای تلفيقي مثل لايدار و يا ساير منابع اطلاعات هندسي موجب افزایش حجم داده و به وجود آمده پدیده کلان داده میشود که پردازش دادهها برای تهیه نقشه دچار مشکلات بسیاری خواهد بود. در این تحقیق، با به کار گیری تلفیقی الگوریتمهای یادگیری عمیق و فیلترهای دو طرفه مسئله استخراج ساختمان از تصاویر نوری با قدرت تفکیک مکانی بالا مورد بررسی و آزمایش قرار گرفت. مدل شبکه عصبی عمیق مورد استفاده در این تحقیق معماری سِگنت میباشد که با استفاده از تابع بهینهسازی برآورد لحظه تطبیقی فرآیند آموزش و تست بهینهسازی گردید و پس از آموزش شبکه عصبی عمیق با استفاده از مجموعه داده آموزشی از شهر پوتسدام آلمان با بهرهگیری از سه بلوک تصویری تست شد. برای ارزیابی دقت نتایج تحقیق از چهار معیار تمامیّت، صحّت، F_{measure} و *IoU* استفاده شد. نتایج حاصل از این ارزیابی نشان داد، به کارگیری فیلتر دو طرفه با پارامترهای بهینهسازی شده تجربی با ابعاد کرنل گوسین ۹×۹ و g_s مقادیر f_r و Λ به ترتیب برای پارامترهای f_r نتایج را تا حد بسیار زیادی بهبود میدهد و مرز ساختمانها به میزان قابل توجهی قابل شناسایی است. به نحوی که برای دادههای تست شماره یک تا سه، مقادیر تمامیت، صحت، F_{measure} و IoU به ترتیب برابر ۹۱٬۴۹، ۹۲٬۳۷، ۹۱٬۴۹ و ۵۷٬۶۷ برای داده تست یک، ۹۱٬۶۷، ۹۰٬۰۲، ۹۰٬۴۴ و ۴۹٬۷۰ برای داده تست دو و ۹۲٬۵۷، ۹۳٬۹۸، ۹۲٬۵۷ و ۵۸٬۳۴ برای داده تست سه است. در بین دادههای تست، داده تست شماره دو،

سال هفتم • شماره دوم • تابستان ۱۳۹۸

کمترین مقادیر دقت و صحت را به خود اختصاص داده است که علت آن به این صورت تحلیل می شود که این منطقه تست، به دلیل وجود ساختمانهای بسیار کوچک و با مرز بسیار کم نسبت به ساختمانهای مجاور به خوبی تشخیص داده نشدهاند، اما پیچیدگیهای هندسی به میزان قابل توجهی در دقت الگوریتم پیشنهادی تحقیق مؤثر نبوده است که این نشان از توان بالای این الگوریتم در حل مسئله استخراج ساختمان دارد.

الگوریتم سِگنت در حل مسئله استخراج اشیا با جزئیات زیاد در مقایسه با معماریهای Pcn-LargeFOV.denseCRF. FCN-DeepLab-LargeFOV-denseCRF.

(learn deconv) و DeconvNet تا حد زیادی موفق عمل کرده است [۴۵]. از طرفی در پژوهشهایی که با هدف استخراج ساختمان صورت گرفته است، از معماری سِگنت بهعنوان یکی از الگوریتمهای مرجع جهت مقایسه نتایج استفاده شده است که معمولاً نتیجه این مقایسهها

مراجع

in Remote Sensing: A Review," IEEE geoscience and remote sensing magazine, in press, Oct. 2017.

- [6] Y. Xu, L. Wu, Z. Xie, and Z. Chen, "Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters," Remote Sensing, vol. 10, no. 1, p. 144, Jan. 2018.
- [7] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 2, pp. 645–657, Feb. 2017.
- [8] M. Radovic, O. Adarkwa, and Q. Wang, "Object Recognition in Aerial Images Using Convolutional Neural Networks," Journal of Imaging, vol. 3, no. 4, p. 21, Jun. 2017.
- [9] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate Object Localization in Remote Sensing Images Based on Convolutional Neural Networks," IEEE Transactions on

در حدود ۵ درصد افزایش دقت کلی گزارش شده است [۶ و ۴۰]. در این پژوهش با بهینه سازی معماری سِگنت با الگوریتم برآورد لحظه تطبیقی و به کارگیری فیلتر دو طرفه، نتایج حاکی از افزایش بیش از ۸ درصد دقت کلی را برای این مجموعه داده می دهد. اما الگوریتم پارامترهای ایکوریتم سِگنت و با استفاده از فیلتر دو طرفه پارامترهای الگوریتم سِگنت و با استفاده از فیلتر دو طرفه بهبود یابد. بعلاوه با مقایسه نتایج روش پیشنهادی تحقیق با روشهای Res-U-Net+GF و الگوریتم پایه در NN+RF+CRF و NSegNet در NN+RF+CRF و NSegNet یاین تحقیق بحث شد، دارای قابلیتهای بهتری نسبت به می توان نتیجه گرفت که الگوریتم SegNet+BF که در این تحقیق بحث شد، دارای قابلیتهای بهتری نسبت به روشهای مشابه در حل مسئله استخراج ساختمان از تصاویر نوری هوایی دارد.

- [1] K. Chen, K. Fu, X. Gao, M. Yan, X. Sun, and H. Zhang, "Building extraction from remote sensing images with deep learning in a supervised manner," 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Jul. 2017.
- [2] Z. Huang, G. Cheng, H. Wang, H. Li, L. Shi, and C. Pan, "Building extraction from multi-source remote sensing images via deep deconvolution neural networks," 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Jul. 2016.
- [3] S. Saito and Y. Aoki, "Building and road detection from large aerial imagery," Image Processing: Machine Vision Applications VIII, Feb. 2015.
- [4] MIT Technology Review, 2018 [Online]. Available: https://www.technologyreview.com/Lists/tec hnologies/2018/.
- [5] X. X. Zhu, D. Tuia, L. Mou, G. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep Learning

مهدی خوش برش ماسوله و همکار ان

Geoscience and Remote Sensing, vol. 55, no. 5, pp. 2486–2498, May 2017.

- [10] M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N. Paragios, "Building detection in very high resolution multispectral data with deep learning features," 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Jul. 2015.
- [11] S. Saito, T. Yamashita, and Y. Aoki, "Multiple Object Extraction from Aerial Imagery with Convolutional Neural Networks," Journal of Imaging Science and Technology, vol. 60, no. 1, pp. 104021– 104029, Jan. 2016.
- [12] Z. Zhong, J. Li, W. Cui, and H. Jiang, "Fully convolutional networks for building and road extraction: Preliminary results," 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Jul. 2016.
- [13] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.
- [14] M. Längkvist, A. Kiselev, M. Alirezaie, and A. Loutfi, "Classification and Segmentation of Satellite Orthoimagery Using Convolutional Neural Networks," Remote Sensing, vol. 8, no. 4, p. 329, Apr. 2016.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84–90, May 2017.
- [16] T. Bluche, H. Ney, and C. Kermorvant, "Feature Extraction with Convolutional Neural Networks for Handwritten Word Recognition," 2013 12th International Conference on Document Analysis and Recognition, Aug. 2013.
- [17] K. Noda, N. Hashimoto, K. Nakadai, and T. Ogata, "Sound source separation for robot audition using deep learning," 2015 IEEE-RAS 15th International Conference on

Humanoid Robots (Humanoids), Nov. 2015.

- [18] G. Fu, C. Liu, R. Zhou, T. Sun, and Q. Zhang, "Classification for High Resolution Remote Sensing Imagery Using a Fully Convolutional Network," Remote Sensing, vol. 9, no. 12, p. 498, May 2017.
- [19] T. Nguyen, J. Han, and D.-C. Park, "Satellite image classification using convolutional learning," American Institute of Physics, pp.2237-2240, Oct 2013.
- [20] Y. Shu, "Deep Convolutional Neural Networks for Object Extraction from High Spatial Resolution Remotely Sensed Imagery," Thesis (PhD), University of Waterloo, 2014.
- [21] F. Alidoost and H. Arefi, "Knowledge Based 3D Building Model Recognition Using Convolutional Neural Networks From LiDAR and Aerial Imageries," ISPRS -International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLI–B3, pp. 833–840, Jun. 2016.
- [22] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by Convolutional Neural Networks," Remote Sensing, vol. 8, no. 7, p. 594, Jul. 2016.
- [23] K. Bittner, S. Cui, and P. Reinartz, "Building Extraction from Remote Sensing Data using fully Convolutional Networks," ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII– 1/W1, pp. 481–486, May 2017.
- [24] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [25] G. Wu, X. Shao, Z. Guo, Q. Chen, W. Yuan, X. Shi, Y. Xu, and R. Shibasaki, "Automatic Building Segmentation of Aerial Imagery Using Multi-Constraint Fully Convolutional Networks," Remote Sensing, vol. 10, no. 3, p. 407, Mar. 2018.

سال هفتم • شماره دوم • تابستان ۱۳۹۸

- [26] G. Goyal, "Impact & Analysis of Improved Bilateral Filter on TEM Images," International Journal of Science and Research (IJSR), vol. 3, Issue 6, Jun. 2014.
- [27] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in International Joint Conference on Neural Networks (IJCNN), 1989.
- [28] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 10, pp. 6232–6251, 2016.
- [29] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data," IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 5, pp. 778–782, May 2017.
- [30] S. Muruganandham, "Semantic Segmentation of Satellite Images using Deep Learning," Thesis (M.Sc), Czech Technical University in Prague, Aug. 2016.
- [31] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network," IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1207– 1216, May 2016.
- [32] D. P. Kingma, and J.L. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv: 1412.6980, 2014.
- [33] H. Adldoost, "Optimization of Image Fusion using Guided Filtering," Thesis (M.Sc), Kharazmi University, Sep. 2015.
- [34] Shutao Li, Xudong Kang, and Jianwen Hu, "Image Fusion With Guided Filtering," IEEE Transactions on Image Processing, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.
- [35] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamicrange images," Proceedings of the 29th annual conference on Computer graphics

and interactive techniques - SIGGRAPH '02, 2002.

- [36] S. Paris, "A gentle introduction to bilateral filtering and its applications," ACM SIGGRAPH 2007 courses on - SIGGRAPH '07, 2007.
- [37] S. Paisitkriangkrai, J. Sherrah, P. Janney, and A. Van-Den Hengel, "Effective semantic pixel labelling with convolutional networks and Conditional Random Fields," 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jun. 2015.
- [38] International Society for Photogrammetry and Remote Sensing (ISPRS) and BSF Swissphoto: WG3 Potsdam overhead data. http://www2.isprs.org/commissions/comm3 /wg4/tests.html.
- [39] P. Kaiser, J. D. Wegner, A. Lucchi, M. Jaggi, T. Hofmann, and K. Schindler, "Learning Aerial Image Segmentation From Online Maps," IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 11, pp. 6054–6068, Nov. 2017.
- [40] N. Audebert, A. Boulch, H. Randrianarivo, B. Le Saux, M. Ferecatu, S. Lefevre, and R. Marlet, "Deep learning for urban remote sensing," 2017 Joint Urban Remote Sensing Event (JURSE), Mar. 2017.
- [41] M. Volpi and D. Tuia, "Dense Semantic Labeling of Subdecimeter Resolution Images With Convolutional Neural Networks," IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 2, pp. 881–893, Feb. 2017.
- [42]https://en.wikipedia.org/wiki/Germay
- [43]https://en.wikipedia.org/wiki/Potsdm
- [44] T. Panboonyuen, P. Vateekul, K. Jitkajornwanich, S. Lawawirojwong, and P. Srestasathiern, "Road Segmentation on Remotely-Sensed Images Using Deep Convolutional Neural Networks with Landscape Metrics and Conditional Random Fields," Jun. 2017.

Downloaded from jgit.kntu.ac.ir on 2025-07-04

مهدی خوش برش ماسوله و همکار ان

[45] V. Badrinarayanan, A. Handa, R. Cipolla, "Segnet: A deep convolutional encoderdecoder architecture for robust semantic pixel-wise labelling," arXiv preprint arXiv: 1505.07293, 2015.



Journal of Geospatial Information Technology Vol.7 No.2, Summer 2019

Research Paper

Integration of Deep Learning Algorithms and Bilateral Filters with Mono Optical Aerial the Purpose of Building Extraction from Imagery

Mahdi Khoshboresh Masouleh¹, Reza Shah-Hosseini^{2*}, Abdol Reza Safari³

PhD Student., School of Surveying & Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran.
 Assistant Professor., School of Surveying & Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran.

3- Professor., School of Surveying & Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran.

Abstract

The problem of extracting the building from mono optical aerial imagery with high spatial resolution is always considered as an important challenge to prepare the maps. The goal of the current research is to take advantage of the semantic segmentation of mono optical aerial imagery to extract the building which is realized based on the combination of deep convolutional neural networks (DCNN) and bilateral filters (BF). For this purpose, considering the hardware limitations of the current research and the fact that it is necessary to select a large number of training data to train deep convolutional neural networks, after selecting an appropriate dataset from three-band optical images, the minimum data that obtains the highest training accuracy was selected to avoid getting weak results due to the lack of training data. In this research, by optimizing the SegNet deep neural network which is an encoderdecoder network, the processing task and therefore extracting the building from optical images are done using the adaptive moment estimation (ADAM) optimization and BF with a Gaussian kernel. This method is implemented on a dataset related to the mono optical aerial imagery of urban regions located in Potsdam, Germany, the twodimensional tagged datasets of international society for photogrammetry and remote sensing (ISPRS). The results show that compared to similar methods, the combinational use of the SegNet optimized deep neural network and BF with a Gaussian kernel provides very appropriate capabilities to improve the detection of building boundary in the optical images with high spatial resolution. Also, the results of the proposed method show that the values of the integrity and validity criteria are 95.14 and 92.37 respectively for the test area 1, 91.67 and 90.2 respectively for the test area 2, and 96.14 and 93.98 respectively for the test area 3.

Key words: building extraction, mono optical aerial imagery, semantic segmentation, deep convolutional neural networks, bilateral filters.

Email: rshahosseini@ut.ac.ir

Correspondence Address. Photogrammetry & Remote Sensing Group, School of Surveying & Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran. Tehran, Tehran, Iran. Tel: +98 21 61114527