نشربه علمى بژو، شي مهندسي فناوري اطلاعات مكاني

سال هفتم، شماره سوم، پاییز ۱۳۹۸ Vol.7, No.3, Autumn 2019 ۱۷۲–۱۹۸ مقاله پژوهشی



# استخراج مدل رقومی سطح با استفاده از تک تصویر ماهوارهای با حدتفکیک بالا و مدل رقومی جهانی SRTM برمبنای یادگیری عمیق

حامد امینی امیرکلائی '، حسین عارفی<sup>\*\*</sup>

۱- دانشجوی دکترای فتوگرامتری دانشکده مهندسی نقشه برداری و اطلاعات مکانی- پردیس دانشکدههای فنی- دانشگاه تهران ۲- استادیار دانشکده مهندسی نقشهبرداری و اطلاعات مکانی، پردیس دانشکدههای فنی، دانشگاه تهران

تاریخ دریافت مقاله: ۱۳۹۷/۱۱/۰۸ تاریخ پذیرش مقاله: ۱۳۹۸/۰۱/۲۱

#### چکیدہ

مدل رقومی سطح (DSM) از جمله مهمترین محصولات در حوزه فتوگرامتری و سنجشازدور میباشد و کاربردهای متنوعی در این حوزه دارد. تکنیکهای موجود به بیش از یک تصویر برای استخراج DSM نیاز دارند و در این مقاله سعی شده است امکان استخراج MSA از تک تصویر ماهوارهای آنالیز و بررسی شود. در این راستا، یک الگوریتم برمبنای شبکههای عصبی کانوولوشنی عمیق طراحی شد. در الگوریتم پیشنهادی ابتدا پیش پردازشهایی نظیر تقسیم تصاویر ماهواره به تصاویر کوچکتر، محلیسازی مقادیر ارتفاعی و تقویت دادههای آموزشی برای آمادهسازی دادهها برای ورود به شبکه انجام میشود. شبکه عصبی کانوولوشنی (CNN) پیشنهادی دارای ساختاری کدگذار کدگشا میباشد که در مرحله کدگذاری ویژگیهای مختلف و کارآمد در مقیاسهای متفاوت استخراجشده و در مرحله کدگشایی و با ارائه روندی کارآمد، ویژگیهای تولیدشده برای تخمین مقادیر ارتفاعی با یکدیگر تلفیق میگردند. سپس با ارائه یک الگوریتم پیکسلهای زمینی و غیرزمینی از هم ارتفاعی ماموریت توپوگرافی رادار شاتل (RTM) با بعاد پیکسل زمینی ۳۰ متر، MSG نهایی بدست میآید. الگوریتم پیشنهادی با استفاده از تفکیکشده و مقادیر ارتفاعی عوارض غیرزمینی استخراج میشوند. با اضافه نمودن عوارض غیرزمینی به همراه اطلاعات ارتفاعی به مدل رقومی ارتفاعی ماموریت توپوگرافی رادار شاتل (RTM) با ابعاد پیکسل زمینی ۳۰ متر، MSG نهایی بدست میآید. الگوریتم پیشنهادی با استفاده از تفکیک شده و مقادیر ارتفاعی عوارض غیرزمینی استخراج میشوند. با اضافه نمودن عوارض غیرزمینی به همراه اطلاعات ارتفاعی به مدل رقومی ارتفاعی ماموریت توپوگرافی رادار شاتل (RTM) با بعاد پیکسل زمینی ۳۰ متر، MSG نهایی بدست میآید. الگوریتم پیشنهادی با استفاده از میتوسط مقادیر ۱۲۹۱۱، ۲۲۱۰ و ۲۹۵۶ مینا در با ای بولی خطای میانگین نسبی (*«R*)، خطای میانگین لگاریتم (*«*A)، ریشسه جذر میانگین مربعات (*«*A)، ۲۲۱۰ و ۲۹۵۶ متر به ترتیب برای خطای میانگین نسبی (*«*A)، خطای میانگین لگاریتم (*«*A)، ریشت برای عربای مربعات (*ه* معادیر (*«*A)، ۲۲۱، و ۲۹۵۶ میای یکیارچه نهایی ایجاد شده به طور متوسط مقدار ۴٫۵۶ متر به ترتیب برای عربی میربی مینگین آمر.

**کلیدواژهها**: مدل سطحی رقومی ، شبکه عصبی کانوولوشنی، تک تصویر ماهوارهای، مدل رقومی ارتفاعی SRTM.

«نویسنده مکاتبه کننده: تهران، امیرآباد شمالی، دانشکده فنی، دانشکده مهندسی نقشهبرداری و اطلاعات مکانی تلفن: ۲۳۰۰۲۵-۳۲۳۰

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

#### ۱– مقدمه

استخراج مقادیر ارتفاعی از تصویر یکی از موضوعات پرچالش و حائز اهمیت است که منجر به تولید مدل سطحی زمین (DSM') به عنوان یکی از محصولات بسیار مهم در فتوگرامتری و سنجش از دور میگردد. DSM را می توان یکی از اجزای مهم جهت انجام پروژههای مختلفی نظیر کلاسهبندی [۱]، قطعهبندی [7]، بازسازی سهبعدی [۳, ۴]، آنالیز صحنه [۵]، شناسایی تغییرات [۶] و به روزرسانی پایگاههای سیستم اطلاعات مکانی [۷] در حوزه فتوگرامتری و سنجش از دور دانست. همانطور که مشخص است، تصاویر رقومی معمولا یک محیط سهبعدی را در دوبعد به نمایش درمی آورند. رایج ترین روش برای محاسبه مقادیر بعد سوم از تصاویر، بهره گیری از تکنیک استریو می باشد که از زوج تصویر اخذ شده در زوایای مختلف از یک صحنه برای انجام مثلثبندی و تعیین موقعیت سهبعدی پیکسلها استفاده مینماید [۸]. گروهی دیگر از روشها نظیر ساختار از حرکت<sup>۳</sup>و شکل از فوكوس<sup>†</sup> از تصاوير تك منظر<sup>6</sup> يا به عبارت ديگر تصاویر اخذ شده از یک موقعیت ثابت برای محاسبه بعد سوم بهره میبرند [۹]. این روشها تک تصویر نبوده و تنها موقعیت تصویر برداری ثابت شده است. در واقع از تصاویر اخذ شده از یک نقطه ثابت که حاوی تغییراتی در موقعیت اشیا و یا فوکوس دوربین میباشند، برای استخراج بعد سوم استفاده می شود. بازیابی اطلاعات سهبعدی تنها با استفاده از تکتصویر فرایندی بسیار پیچیده و از لحاظ ریاضیاتی بدوضع<sup>۶</sup> است[۱۰، ۱۱، ۱۲ و ۱۳]. درواقع تصویر کردن مقادیر درجه خاکستری یا رنگی به اطلاعات سهبعدی دارای یک ابهام ذاتی است؛ چرا که هیچ رابطه مستدل و دقیقی بین دادههای رنگ و شدت هر پیکسل و ارتفاع آن از تصویر وجود ندارد [۱۴]. برای مثال تعدادی عارضه با شکل یکسان و اندازههای مختلف که در

- <sup>3</sup> Structure from motion
- <sup>4</sup> Shape from focusing
   <sup>5</sup> Single view
- <sup>6</sup> Ill-pose

فواصل متفاوتی نسبت به یک مشاهده کننده قرار دارند، می توانند تصویر یکسانی برای مشاهده کننده ایجاد نمایند. روش شکل از سایه روشن<sup>۷</sup> را می توان به عنوان یک روش استخراج بعد سوم از تک تصویر در نظر گرفت که دارای عملکردی قابل قبول است [۱۵]. این روش به صورت محدود برای استخراج مقادیر ارتفاعی از تصاویر ماهوارهای نیز استفاده شده است، اما ایشتر برای استخراج مدلهای رقومی کلی برای مناطق وسیع با دقت کم کارایی داشته است [۱۶ و مناطق وسیع با دقت کم کارایی داشته است [۱۷ ینلر گرفتن ویژگیهای دیگر موجود در تصویر نفی تواند برای دستیابی به مدلهای سه بعدی دقیق مناسب باشد.

انسان به صورت ذاتی از دوچشم جهت اخذ آنی تصویر و تعیین فاصله بهره میبرد. درعین حال از توانایی تخمین ساختار محیط حتی از طریق دید تکچشمی و با استفاده از تک عکس نیز بهرهمند است که ناشی از آموزش مغز در طول زمان و با بررسی و استنتاج اطلاعات مختلف تصويري ميباشد. ازاينرو جهت حل مساله تخمین عمق از تک تصویر در حوزه پردازش تصویر و بهصورت محاسباتی نیاز است که نحوه عملكرد سيستم بينايي انسان بهصورت دقيق بررسي شده و دانش حاصل از این بررسی ها با استفاده از سختافزارها و نرمافزارهای کامپیوتری پیادهسازی شود[۱۸]. قشر بصری چشم انسان از المانهای هادی^ در دید استریو و تکچشمی<sup>۹</sup> برای تخمین عمق نسبی صحنه استفاده مینماید. درواقع زمانی که دید استریو در دسترس نباشد (برای مثال هنگام مشاهده یک تصویر و یا دچار نقص بینایی شدن) مستقیماً از المانهای هادی تکچشمی برای استنتاج و تفسیر عمق از صحنه استفاده می شود [۱۹]. به طور کلی، در مغز انسان ساختار و هندسه یک صحنه از طریق تحلیل ویژگیهای آن صحنه نظیر عوارض خطی، اندازه نسبی اشیا، سایه و غیره در راستای پاسخ به این سؤال که کدام عارضه نزدیکتر بوده و یا در جلوی

<sup>&</sup>lt;sup>1</sup> Digital Surface Model (DSM)

<sup>&</sup>lt;sup>2</sup> Change detection

<sup>&</sup>lt;sup>7</sup> Shape from shading

<sup>&</sup>lt;sup>8</sup> Cue

<sup>&</sup>lt;sup>9</sup> Monocular

دیگر عوارض قرار گرفته است، استنتاج می گردد. بازسازی ساختار هندسی یک صحنه از تک تصویر نیز بسیار پرکاربرد بوده و میتواند در زمینههای مختلفی نظیر شناسایی مرز نواحی پنهان، قطعهبندی صفحات ارتفاعي و استفاده بهعنوان قدم اول جهت دستيابي به نقشه كامل ارتفاعى مفيد فايده باشد. بهطوركلى، استفاده از ویژگیهای خاص و یا محلی موجود در یک تصویر جهت تخمین بعد سوم بسیار ناکافی است و در کنار اطلاعات محلی نیاز به داشتن اطلاعاتی از ساختار کلی تصویر است. شبکههای عصبی کانوولوشنی (CNN)<sup>۱</sup> بهعنوان گروهی از الگوریتمهای یادگیری عمیق<sup>۲</sup>، دارای توانایی مهندسی ویژگی<sup>۳</sup> و استخراج اطلاعات در سطوح و مقیاسهای مختلف از تصویر مى باشند [70]. بەطورى كە بەتدرىج ابعاد لايەھاى ویژگی کوچک شده و تعداد لایههای ویژگی استخراجشده افزایش می یابد [۲۰]. ایجاد یک ارتباط متوازن و دقیق بین خصوصیات محلی استخراجشده از لایههای کمعمق<sup>†</sup> و خصوصیات کلی استخراجشده از لايەھاى عميق<sup>4</sup> يک صحنه مىتواند به بازسازى ھرچە بهتر مدل سهبعدی آن صحنه کمک شایان توجهی نمايد [٢١].

ماموریت توپوگرافی رادار شاتل<sup>۶</sup> (SRTM) یک پروژه تحقیقاتی جهانی در سال ۲۰۰۰ برای بدست آوردن مدل ارتفاعی رقومی<sup>۷</sup> (DEM) از سطح زمین با دقت مناسب بین مدارهای ۶۰ درجه شمای و ۵۶ درجه جنوبی است. قدرت تفکیک مکانی مدلهای ارتفاعی SRTM، یک ثانیه قوسی یا ۳۰ متر است که ابتدا تنها امریکا به این قدرت تفکیک دسترسی داشت، اما امروزه امکان دانلود دادههای ارتفاعی با دقت ۳۰ متر برای همه قابل دسترس است. به طورکلی، قدرت تفکیک ۳۰ متر میتواند یک ساختار کلی از ناهمواریهای زمین در اختیار کاربر قرار داده و امکان

- <sup>1</sup> Convolutional neural network
- <sup>2</sup> Deep learning

<sup>6</sup> Shuttle Radar Topography Mission

بررسی تغییرات ارتفاعی ناشی از عوارضی نظیر ساختمانها و درختان را فراهم نمی سازد. در تحقیق حاضر تلاش شده است که با ارائه یک شبکه عمیق قدرتمند، یک DSM با جزئیات بالا تنها با از تکتصویر ماهوارهای تخمین زده شود. در این راستا، ابتدا با استخراج ویژگیهای سطح بالا<sup>۸</sup>، ساختار کلی منطقه بازسازی گشته و سپس با استفاده از ویژگیهای سطح پایین<sup>۹</sup>، مقادیر ارتفاعی عوارض کوچکتر موجود و جزئیات در منطقه مطالعاتی تخمین زده می شوند. لازم به ذکر است که ویژگیهای سطح پایین به ویژگیهای استخراج شده از لایههای اولیه و کمعمق و ویژگیهای سطح بالا به ویژگیهای استخراج شده از لایههای عمیق در یک شبکه CNN گفته می شود. در نهایت با ارائه روندی کارآمد عوارض غیرزمینی و اطلاعات ارتفاعی آنها استخراج شده و با بهره گیری از مدل رقومی ارتفاعی SRTM، یک DSM

یکپارچه و دقیق برای تصویر ماهوارهای بدست آید. در ادامه ساختار مقاله به این صورت است که در بخش ۲ مروری بر مطالعات انجام شده در حوزه استخراج تخمین مقادیر ارتفاعی و عمق از تک تصویر ارائه شده است. سپس در بخش ۳ شرح مختصری از ساختار شبکههای عمیق CNN بیان شده و سپس روش پیشنهادی بهطور کامل در بخش ۴ شرح داده شده است. در بخش۵، روند پیشنهادی پیادهسازی گشته و مورد ارزیابی قرار می گیرد و درنهایت در بخش ۶ نتیجه گیری و پیشنهادات آتی ارائه می شوند. ۲ – پیشنیه تحقیق

تحقیقات صورت گرفته در زمینه استخراج بعد سوم را می توان در دو گروه روشهای استخراج عمق از تک تصویر که مربوط به تصاویر در حوزه بینایی ماشین هستند و روشهای تخمین ارتفاع از تک تصویر که مربوط به تصاویر در حوزه فتو گرامتری و سنجش از دور هستند، تقسیم نمود.

### ۲–۱– تخمین مقادیر عمق از تک تصویر

روشهای استخراج بعد سوم عمدتاً در حوزه بینایی ماشین و جهت تخمین عمق از تصاویر اخذشده در

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

<sup>&</sup>lt;sup>3</sup> Feature engineering

<sup>&</sup>lt;sup>4</sup> Shallow layers

<sup>&</sup>lt;sup>5</sup> Deep layers

<sup>&</sup>lt;sup>7</sup> Digital Elevation Model

<sup>&</sup>lt;sup>8</sup> High-level

<sup>&</sup>lt;sup>9</sup> Low-level

محیطهای شهری بوده است. این روشها را میتوان در دو گروه روشهای مبتنی بر مدلهای گرافیکی و روشهای مبتنی بر یادگیری عمیق تقسیم نمود. ۲-۱-۱- روشهای مبتنی بر مدلهای گرافیکی در این روشها تولید ویژگیهای مناسب نقش مهمی در نتایج حاصله دارد. در این روشها با ایجاد یک مجموعه داده آموزشی براساس ویژگیهای مختلف تصویر و بهره گیری از الگوریتمهای میدان تصادفی شرطی و مارکفی مقادیر عمق از تصاویر رنگی استخراج میشوند. محققان روشهای مختلفی را برای استخراج ویژگی از تکتصویر و استفاده از آن در تخمین مقادیر عمق مطرح نمودهاند. به طوری که، ساکسنا و همکاران (۲۰۰۶)، از یک روش نظارتشده برای تخمین عمق استفاده نمودند. بهطوریکه، ابتدا تصویر به قطعات کوچک تقسیم شده و برای هر قطعه یک عمق محاسبه گردید. برای استخراج ویژگی در هر مقياس از قطعات مجاور قطعه موردنظر نيز استفاده شد که این کار سبب در نظر گرفتن همسایگی نزدیک و همسایگی دور گردید. درنهایت برای هر قطعه یک بردار ویژگی تشکیل شد که حاوی اطلاعات کارآمدی برای تخمین عمق بود. از دو مدل احتمال گوسین و لاپلاسین میدان تصادفی مارکفی برای مدلسازی با بهرهگیری از بردارهای ویژگی تشکیل شده، استفاده شد که مدل لایلاسین جواب بهتری داشت [۲۲, ۲۲]. ليو و همكاران (۲۰۱۴)، مساله استخراج عمق از تک تصویر را بهصورت یک مساله بهینهسازی گسسته-پیوسته درنظر گرفتند و برای حل آن از الگوریتم میدان تصادفی شرطی استفاده نمودند. بهطوریکه متغیرهای پیوسته عمق را در سوپرپیکسلها محاسبه کرده و متغیرهای گسسته رابطه بین سوپرپیکسلهای همسایه را نشان دادند [۲۳]. ژو و همکاران (۲۰۱۵) یک مدل سلسله مراتبی برای استخراج عمق مطرح نمود. در این مدل از الگوریتم میدان تصادفی شرطی جهت تعیین رابطه بین لایههای مختلف در طی سلسلهمراتب استفاده شد. منظور از تخمین عمق

سلسلهمراتبی، استخراج عمق در سه سطح محلی، میانی و کلی بوده است.

### ۲-۱-۲ روشهای مبتنی بر یادگیری عمیق

در این روشها به جای تولید ویژگی با شیوههای مختلف و به صورت دستی، با ارائه ساختارهای متنوع سعی شده است که ویژگیهای مناسب و کارآمد در داخل شبکه تولید شوند. اساس روشهای مبتنی بر یادگیری عمیق طراحی یک شبکه کانوولوشنی و قراردادن لایههای مختلف شبکه جهت دستیابی به بهترین دقت میّباشد. در این راستا، روشهای مختلفی ارائه شده است که در ادامه به صورت مختصر به آنها پرداخته می شود. ایجن و همکاران (۲۰۱۴)، روشی شامل دو بخش اصلی پیشنهاد نمودند. بهطوریکه ابتدا با استفاده از یک شبکه اولیه یک پیشبینی کلی از عمق صحنه بدست آمده و سپس با استفاده از یک شبكه بهبوددهنده مقادير عمق بهصورت محلى بهبود داده شدند [11]. هر دو شبکه از تصویر اولیه بهعنوان ورودى استفاده نموده و خروجي تخمين اوليه بهعنوان ورودي تخمين دقيق نيز استفاده گشت. هدف طراحي شبکه بهبوددهنده، تنظیم پارامترهای تخمین اولیه با جزئیات محلی بوده است که این کار با طراحی شبکهای با لایههای کانوولوشن و یک لایه ادغام صورت پذیرفت. ژانگ و همکاران (۲۰۱۸)، یک شبکه سلسله مراتبی ارائه نمودند که در آن سعی شد ویژگیهای محلی و جزئیات موجود در صحنه بهصورت تدریجی در روند تخمین عمق و طی گذر از لایههای کدگشایی<sup>۳</sup> بازیابی شوند. در این راستا، از المانهای مختلف هندسی در سطوح مختلف شبکه جهت تنظیم پارامترها طی روند آموزش استفاده شده است. از نرم L2 بهعنوان المان هندسی در سطح پیکسل، از نرم L2 گرادیان، نرم L2 گرادیان و گرادیان عملگر سوبل به عنوان المان هندسی در سطح ناحیه و از کوواریانس دوطرفه به عنوان تنظیم کننده ثبات کلی در شبکه استفاده شده است [۲۱]. هی و وانگ (۲۰۱۸) سعی نمودند که ابهام بین فاصله کانونی و تخمین عمق از تکتصویر را حل نمایند. در این راستا، الگوریتمی جهت تولید تصاویری با فاصله کانونیهای متفاوت با

<sup>3</sup> Decoding

<sup>&</sup>lt;sup>1</sup> Conditional random field

<sup>&</sup>lt;sup>2</sup> Markov random field

<sup>[</sup> Downloaded from jgit.kntu.ac.ir on 2025-07-04

استفاده از تصاویر با فاصله کانونی ثابت پیشنهاد نمودند. سپس یک شبکه کدگذار-کدگشا با توانایی آموزش جهت تخمین عمق با رزولوشن مناسب از تک تصوير از طريق تلفيق اطلاعات سطوح مياني را ارائه نمودند. درنهایت با بهره گیری از اطلاعات فاصله کانونی مربوط به تصاوير توليدشده با فواصل كانونى متغير جهت بهبود تصاوير عمق تخمين زدهشده، استفاده شد. لی و همکاران (۲۰۱۸)، یک چهارچوب عمیق بر اساس تورم کانوولوشن و استفاده از ساختار شبکه ResNet ارائه نمود که توانایی توزیع احتمال در میان برچسبهای عمق را دارد. در شبکه مذکور برای كاهش تعداد پارامترها، لايههاى تماماً متصل حذف گشتند. ایراتور تورم کانوولوشن امکان گسترش ناحیه یذیر ش<sup>۱</sup> را بدون افزایش تعداد پارامترها فراهم مىنمايد. بەعلاوە با استفادە از اين اپراتور نياز بە اپراتور ادغام<sup>۲</sup> بدون کاهش ابعاد زمینه پذیرش و از دست دادن رزولوشن مکانی، از بین میرود. برای استفاده از اطلاعات محلی نقشههای ویژگی لایههای میانی، این لایهها مستقیماً به لایه نهایی نقشه ویژگی الحاق شدند [۲۴].

### ۲-۲- تخمین مقادیر ارتفاع از تک تصویر

در چند سال اخیر تحقیقات محدودی در زمینه استخراج مقادیر ارتفاعی از تکتصویر در حوزه فتوگرامتری و سنجشازدور صورت گرفته است که تمام آنها به دلیل قدرت بالای الگوریتمهای یادگیری عمیق از این الگوریتمها در این زمینه استفاده نمودهاند. سریواستاوا و همکاران(۲۰۱۷)، از یک ساختار CNN برای کلاسهبندی و همین طور تخمین مدل رقومی سطحی نرمال شده (۳*MSM*) از تصاویر هوایی استفاده نمود [۲۵]. مو و همکاران (۲۰۱۸)، یک شبکه برای تخمین مقادیر ارتفاعی از تصاویر سنجش زدور ارائه نمودند. در این روش از *MSM* حاصل از سنجنده لایدار و تصاویر سنجش ازدور جهت آموزش شبکه استفاده شده است. همچنین از یک

177

كدگذارى<sup>4</sup> به آخرين لايه كدگشايى جهت بهبود

عملکرد شبکه در لبه ساختمانها بهره گرفته شد [۲۶]. قمیسی و همکاران (۲۰۱۸)، از یک شبکه عمیق مولد شرطی تقابلی<sup>۶</sup> که دارای یک معماری کدگذار-کدگشا به همراه اتصالات جهشی است برای شبیهسازی مقادیر ارتفاعی از تصاویر هوایی بهره برد. در ارزیابیها مشخص شد که استفاده از DSM تولید شده در این روش می تواند در افزایش دقت كلاسەبندى عوارض بسيار كمككنندە باشد [٢٧]. امینیامیرکلائی و عارفی (۲۰۱۹)، روندی برمبنای یادگیری عمیق برای استخراج مقادیر ارتفاعی از تکتصویر هوایی ارائه دادند که در آن ابتدا طی یک روند پیش پردازش، لبههای موجود در تصویر تقویت شدند. سیس یک شبکه CNN با ساختاری U ارائه گشت که در آن با بهرهگیری از اتصالهای جهشی سعی شد که از جزئیات موجود در لایههای کمعمق در روند تخمين ارتفاع دوباره استفاده شود. سپس با آموزش شبکه CNN مطرح شده، مقادیر ارتفاعی تصاویر ورودی محاسیه گشت. در نهایت با ارائه یک روند پس پردازش، تصاویر ارتفاعی تخمین زده شده به یکدیگر متصل شده و یک سطح ارتفاعی پیوسته ایجاد شد [۲۸]. امینی امیر کلائی و عارفی (۲۰۱۹)، با تغییر و بهبود روند اتصال تصاویر ارتفاعی تخمین زده شده (در مقایسه با روش مطرح شده در [۲۸]) سعی نمودند که مدل های ارتفاعی نهایی ایجاد شده، به شکل بهتری ناهمواریهای سطح زمین را نمایش دهند. همچنین آنها با بررسی سناریوهای متفاوت در پیادهسازی و بهره گیری از تصاویر متنوع زمینی، هوایی و ماهوارهای عملکرد شبکه را در شرایط مختلف آنالیز

### ۲-۳- نوآوری تحقیق

نمودند [۲۹].

در تحقیق پیشرو نیز با ارائه یک شبکه عمیق CNN به تخمین مقادیر ارتفاعی پرداخته شده است. در شبکه پیشنهادشده، برخلاف برخی شبکهها که تنها در لایه آخر از اتصالهای جهشی استفاده نمودهاند [۲۶]، در تمامی سطوح کدگشایی استفاده بهره گرفته شد.

<sup>&</sup>lt;sup>1</sup> Receptive field

<sup>&</sup>lt;sup>2</sup> Pooling

<sup>&</sup>lt;sup>3</sup> normalize Digital Surface Model

<sup>&</sup>lt;sup>4</sup> Skip connection

<sup>&</sup>lt;sup>5</sup> Encoding

<sup>&</sup>lt;sup>6</sup> Conditional generative adversarial

همچنین بهجای استفاده از لایههای درونیابی در روند كدگشایی كه سبب افزایش تعداد یارامترها و كاهش سرعت آموزش می گردد [۲۶، ۲۸ و ۲۹]، از اعمال چند کانوولوشن با ابعاد متفاوت و سپس تجمیع آنها بهره گرفته شد. در شبکههای مطرح شده در [۸۲ و ٢٩]، الگوريتمهايي براي اتصال تصاوير ارتفاعي تخمين زده شده و ایجاد مدل ارتفاعی پیوسته پیشنهاد شده است که این روشها عمدتاً در مناطق کوچک با شیب ملايم عملكرد مناسبي داشته و در صورت وسعت بالا و ناهموارى زياد منطقه مطالعاتى عملكرد ضعيفي دارند. چراکه محدوده کوچک تصاویر موردپذیرش در شبکههای CNN سبب می گردد که شیب سطح زمین در تصاویر مذکور به خوبی در شبکه مدل سازی نشود و شبکه نسبت به شیب کلی منطقه به درستی آموزش نبیند. بهعبارتدیگر در محدودههای کوچک شیب کلی سطح زمین قابل تشخیص و مدلسازی دقیق نمیباشد. ازاینرو در این مقاله با ارائه رویکردی متفاوت و بهره گیری از مدل رقومی جهانی SRTM و همین طور ارائه روشی برای تفکیک نقاط زمینی و غیرزمینی در تصاویر تخمین زده شده سعی شد تا روندى كارآمد براى ايجاد مدلهاى رقومى ارتفاعى پیوسته حتی در صورت وسعت زیاد و ناهمواری شدید منطقه ارائه شود.

### ۳- شبکههای عمیق CNN

شبکههای *CNN* از مهمترین و رایجترین روشهای یادگیری عمیق هستند و دارای الگوریتمهای آموزشی پایدار میباشند که امکان یادگیری نمایشهای تصویری را بدون نیاز به طراحی و تولید دستی ویژگیها فراهم مینمایند. عملکرد این شبکهها همانند تمام الگوریتمهای شناسایی بسیار وابسته به دادههای آموزشی است؛ اما به دلیل عملکرد موفق آنها در مقابل دادههای حجیم و بزرگمقیاس نسبت به دیگر الگوریتمها بسیار موردتوجه کارشناسان قرارگرفته است [۳۰]. در هر شبکه *CNN* دو مرحله پیشرونده<sup>۱</sup> و پسانتشار<sup>۲</sup> برای آموزش وجود دارد. در مرحله اول

چیزی جز ضرب نقطهای بین ورودی و پارامترهای هر نورون و نهایتاً اعمال عملیات کانوولوشن در هر لایه نیست و تا خروجی شبکه محاسبه شود. بهمنظور تنظیم پارامترهای شبکه، خروجی شبکه با استفاده از یک تابع اتلاف با داده مرجع مقایسه گشته و بدین ترتیب میزان خطا محاسبه می شود. در مرحله بعد بر اساس میزان خطای محاسبه شده مرحله پس انتشار آغاز می شود که در آن گرادیانت هر پارامتر با توجه به قانون زنجیره ای محاسبه شده و تمامی پارامترها با توجه به تأثیری که بر خطای ایجاد شده در شبکه دارند، تغییر می یابند. پس از چند تکرار جواب نهایی شبکه بدست آمده و شبکه پایان می یابد. در ادامه مؤلفه های اصلی CNN به صورت خلاصه توضیح داده شده اند.

**لایه کانوولوشن:** لایه کانوولوشن هسته اصلی تشکیل دهنده *CNN* است و هر لایه آن از تعدادی فیلتر یا کرنل ساخته شده که شامل وزنها و یک بایاس میباشد. این عملگر اطلاعات و ویژگیهای مختلفی از داده ورودی که میتواند تصویر ورودی و یا خروجی لایه قبل باشد، استخراج مینماید. مکانیزم اشتراک وزن در عملگرهای کانوولوشن تعداد پارامترها

را کاهش و سرعت آموزش را افزایش می دهد [۳۱]. **لایه ادغام:** این لایه معمولاً بعد از لایه کانوولوشن قرار می گیرد و از آن برای تغییر اندازه نقشههای ویژگی استفاده می شود. استفاده از تابع ماکزیمم در این لایهها سبب همگرایی سریعتر، تعمیم بهتر و انتخاب ویژگیهای نامتغیر بسیار عالی می شود [۳۲]. لایه تماماً متصل<sup>۳</sup>: این لایهها معمولاً آخرین لایههای یک شبکه را تشکیل می دهند که یک نقشه ویژگی را به یک بردار ویژگی تبدیل می نماید. لایههای تماماً متصل همانند همتایان خود در شبکههای عصبی مصنوعی سنتی عمل کرده و تقریباً.<sup>(۹</sup> پارامترهای یک شبکه *CNN* را شامل می شوند [۳۳]. مشکل بزرگ این نوع لایهها این است که دارای تعداد بسیار را بالایی نیاز دارد [۳۳].

<sup>&</sup>lt;sup>1</sup> Feed forward

<sup>&</sup>lt;sup>2</sup> Back propagation

<sup>&</sup>lt;sup>3</sup> Fully connected

**توابع غیرخطی**: توابع غیرخطی بهمنظور مدلسازی روند فعالسازی نورنهایی خاص و جدا نمودن دادههایی است که بهصورت خطی قابل جداسازی نمیباشند. توابع سیگموئید، تانژانت هیپربولیک و واحد خطی اصلاح شده (*ReLU*<sup>1</sup>) ازجمله توابع رایج در این زمینه میباشند که تابع *ReLU کار*آمدترین تابع در شبکههای *CNN* میباشد [۳۴]. نحوه عملکرد این تابع به این صورت است که با اعمال آن مقادیر منفی مفر شده و مقادیر مثبت را بدون تغییر نگه داشته میشوند. ساختار ساده آن سبب کاهش هزینه میشود [۳۵].

**لایهنرمالسازی دستها ی<sup>7</sup>:** این لایهها سبب یکسانسازی محدوده دادهها و افزایش سرعت پردازش می گردد. نرمالسازی دستهای سبب می شود مقدار شیفت کوواریانس کاهش یابد و هر لایه مقداری مستقل تر عمل نماید. شیفت کوواریانس زمانی به و چود می آید که شبکه با مقادیر x و y آموزش می بیند و پراکندگی x تغییر می یابد و برای اصلاح آن باید از پراکندگی y استفاده نمود. نرمالسازی دادههای با اطمینان حاصل کردن از این که هیچ مقداری در تابع فعال سازی مقداری بسیار بزرگ یا بسیار کوچک ندارد، سبب می شود که بتوان مقادیر بزرگ تری برای نرخ آموزش<sup>۳</sup> در نظر گرفت که منجر به افزایش سرعت همگرایی می گردد [۳۶].

**لایه حذف تصادفی<sup>3</sup>:** لایه حذف تصادفی به منظور جلوگیری از بیش برازش<sup>6</sup></sup> در شبکه های عصبی معرفیشد و نحوه کار آن به این صورت است که در هرمرحله از آموزش، هر نورون یا با احتمال <math>1-q (از شبکه) بیرون انداخته شده و یا با احتمال q نگه داشته میشود، به طوری که نهایتاً یک شبکه کاهش داده شده باقی بماند [۳۷]. لازم به ذکر است که شبکه های *CNN* می توانند تنها دارای ساختار کدگذاری باشند و یا علاوه بر آن دارای ساختار کدگشایی نیز باشند.</sup></sup>

در مرحله کدگذاری بهتدریج ابعاد ماتریس نقشههای ویژگی تولیدشده کاهشیافته و تعداد ویژگیهای توليدشده (عمق ماتريس) افزايش مىيابد. افزايش عمق در شبکههای CNN با قرار دادن لایههای کانوولوشن صورت می پذیرد. اما، بدون در نظر گرفتن راهکاری مناسب و تنها با قرار دادن لایهها بهصورت پیاپی، شبکه با مشکل صفر شدن گرادیان<sup>۶</sup> در مرحله پسانتشار و عدم افزایش دقت شبکه مواجه می گردد. در معماری ResNet، نگاشت باقیمانده جهت حل مساله صفر شدن تدريجي گراديان ارائه شده است. . اگر نگاشت (H(x) در نظر گرفته شود، چند لایه پیایی یک نگاشت F(x) = H(x) - x انجام داده و درنهایت با بهره گیری یک اتصال جهشی نتایج بدست آمده با مقدار x جمع می گردند (F(x) + x). به عبارت دیگر در نگاشت باقیمانده، مقادیر خروجی پیش از نگاشت با نتایج خروجی نگاشت جمع می گردند [۳۸]. به این ترتيب امكان افزايش عمق و همراه افزايش دقت فراهم می شود. مرحله کدگذاری به تعدادی لایه تماما متصل منتهی می شود که تعداد آن ها متناسب با اهداف مورد نظر جهت شناسایی تعیین می شوند. در نتیجه می توان گفت که شبکههایی که تنها دارای مرحله کدگذاری هستند، برای اهدافی نظیر قطعهبندی و کلاسهبندی که منطقه را به تعدادی کلاس محدود تقسیم مى نمايند، مناسب اند.

اهدافی نظیر تخمین مقادیر عمق و ارتفاعی که هر پیکسل مقداری مجزا داشته و هدف به تعدادی کلاس محدود نمی شود، از ساختار کدگشایی در شبکه استفاده می شود که در آن از روش هایی برای افزایش ابعاد نقشه های ویژگی استفاده می شود. از روش های رایج در این زمینه می توان به عملگرهای معکوس کانوولوشن و ادغام اشاره کرد [۳۹و۴۰]. در این روش ها مناسبی درنهایت حاصل نمی شود. به طوری که در مناسبی درنهایت حاصل نمی شود. به طوری که در شبکه برای دستیابی به جوابی مناسب دشوار است و شببه بابجایی در موقعیت و همین طور ساختار هندسی عوارض در تصویر می شود. در روش معکوس

<sup>&</sup>lt;sup>1</sup> Rectified Linear Unit

<sup>&</sup>lt;sup>2</sup> Batch normalization

<sup>&</sup>lt;sup>3</sup> Learning rate

<sup>&</sup>lt;sup>4</sup> Drop out

<sup>&</sup>lt;sup>5</sup> Overfitting

<sup>&</sup>lt;sup>6</sup> Vanishing gradient

۴- روش پیشنهادی

در این بخش روند پیشنهادی برای استخراج DSM از

تک تصویر ماهوارهای به صورت کامل شرح داده شده

است. در شکل (۱) مراحل روند پیشنهادی نشان داده

شده است. ساختار کلی روش پیشنهادی به این صورت است که ابتدا یردازشهایی برای آمادهسازی و تقویت

دادههای آموزشی صورت می پذیرد. سپس از دادههای

آموزشی برای آموزش شبکه CNN طراحی شده، استفاده می گردد. در مرحله بعد شبکه کانوولوشنی

آموزش دیده برای تخمین تصاویر ارتفاعی اولیه به کار

گرفته شده و با ارائه یک الگوریتم، پیکسلهای زمینی

و غیرزمینی در تصاویر ارتفاعی تخمین زده شده از

یکدیگر تفکیک می شوند. در نهایت با اضافه نمودن

مقادیر ارتفاعی پیکسلهای غیرزمینی در تصاویر

ادغام، پس از دوبرابر شدن ابعاد ماتریس ورودی پیکسلهای بدست آمده از تصویر بهصورت قراردادی مقدار گرفته و مابقی صفر میشوند [۴۱]. درواقع حدود ۷۵ درصد از پیکسلهای بهوجود آمده صفر هستند که نشان از بیهوده بودن حجم زیادی از محاسبات میباشد که برای ازبین بردن مقادیر از عملگر کانوولوشن و اعمال آن به تصویر استفاده میشود. برای جلوگیری از انتخاب قراردادی پیکسل میشود. برای مقداردهی، از ساختارهای متقارن در شبکههای عملگر ادغام در مرحله کدگذاری یک عملگر معکوس ادغام در مرحله کدگشایی استفاده میشود. مکانهایی که طی اعمال عملگر ادغام انتخاب شدهاند، در مرحله معکوس ادغام برای مقداردهی استفاده میشوند [۳۹].



شکل ۱: روندکلی الگوریتم پیشنهادی

در راستای شرح کامل روند پیشنهادی، در بخش ۴–۱، روندی برای پیش پردازش و آمادهسازی دادهها برای ورود به شبکه ارائه می شود. در بخش ۴–۲، اجزای شبکه CNN پیشنهادی برای استخراج مقادیر ارتفاعی از تک تصویر به طور کامل توضیح داده می شود. درنهایت در بخش ۴–۳، روند پیشنهادی برای ایجاد DSM نهایی تشریح شده است.

۴-۱- پیش پردازش و آمادهسازی دادهها

از آنجاکه تصاویر فتو گرامتری و سنجش ازدور معمولاً منطقه وسیعی را پوشش داده و ابعادی بزرگی دارند، نمی توان این ابعاد از تصویر را به صورت مستقیم به شبکه های CNN وارد کرد. از این رو این تصاویر باید مطابق ابعاد ورودی مورد قبول در شبکه درون یابی شده

استخراج مدل رقومی سطح با استفاده از تکتصویر ... حامد امینی امیرکلائی ، حسین عار فی

و سپس وارد شبکه شوند. از طرفی، اگر تصویر کل منطقه مطالعاتی با انجام درونیابی به ابعاد مورد پذیرش در شبکه تبدیل شوند، اطلاعات زیادی از دست رفته و تصویر نامفهومی حاصل میشود که امکان بازیابی مقادیر ارتفاعی را از بین می برد. درنتیجه تصاویر ماهوارهای باید به تصاویر کوچکتری تقسیم شوند. ماهوارهای باید به تصاویر کوچکتری تقسیم شوند. به عبارت دیگر، به ازای هر تصویر متناسب با ابعاد آن، به عبارت دیگر، به ازای هر تصویر متناسب با ابعاد آن، تعدادی تصویر کوچکتر ایجاد گشته و سپس با بعدادی تصاویری با ابعاد مورد پذیرش در شبکه *د*رونیابی، تصاویری با ابعاد مورد پذیرش در شبکه درونیابی می شوند. در شکل(۲) نحوه تقسیم تصویر مطابق با ابعاد موردقبول در شبکه آورده شده است.



شکل ۲: روند تقسیم تصاویر رقومی که محدوده زیادی را پوشش میدهند به تصاویر کوچک تر متناسب با ابعاد موردقبول در شبکه CNN

استفاده از دادههای ارتفاعی با محدوده متغیر سبب عدم همگرایی مناسب شبکه می شود. به طوری که شبکه آموزش دیده نمی تواند مقادیر ارتفاعی را برای مناطقی که دارای محدوده ارتفاعی متفاوتی نسبت به دادههای استفاده شده در روند آموزش می باشند را به خوبی تخمین بزند. برای مثال، دو ساختمان یکسان در دو منطقه با ارتفاعهای ۲۰۰۰ متر و ۱۰ متر نسبت به سطح دریا، می توانند سبب تغییرات مقادیر ارتفاعی یکسانی در محیط باشند، اما شبکه عملکرد یکسانی در قبال آن ها ندارد. درواقع شبکه آموزش دیده با استفاده از دادههای ارتفاعی منطقه با ارتفاع ۲۰۰۰ متر، دیگر قابلت تخمین ارتفاع در منطقه با ارتفاع ۲۰۰۰ متر، دیگر

ندارد. درنتیجه میتوان گفت که میزان تعمیم پذیری شبکه کاهش مییابد. جهت حل این مسأله مقادیر ارتفاعی در هر تصویر بریدهشده با کسر از کمینه ارتفاع در آن منطقه بهصورت محلی درآمد. این کار سبب میشود که مقادیر ارتفاعی ساختمانهایی دارای ساختار یکسان بوده و تغییرات ارتفاعی یکسانی در منطقه ایجاد مینمایند، بهخوبی توسط شبکه شناساییشده و مقادیر ارتفاعی آنها تخمین زده شوند. شناساییشده و مقادیر ارتفاعی آنها تخمین زده شوند. که آموزش آن نیاز به مجموعه بزرگی از دادههای آموزشی دارد تا مساله واگرا نشود. اما در اغلب موارد دادههای آموزشی محدود بوده و برای جلوگیری از واگرا

کدگشا<sup>۲</sup> است که پس از استخراج ویژگی در روند شدن شبکه از روشهای تقویت داده<sup>۱</sup> استفاده میشود. کدگذاری، یک روند کدگشایی برای بزرگ کردن ابعاد نقشههای ویژگی و بهتدریج تبدیل آنها به تصویر ارتفاعی را شامل می شوند. در شکل (۳) ساختار شبکه پیشنهادی نشان داده شده است. ۴-۲-۱- مرحله کدگذاری در مرحله کدگذاری به دلیل قدرت بالای شبکه عمیق باقیمانده (ResNet) از ساختار این شبکه بهره گرفته شده است. ساختار به کاررفته در مرحله کدگذاری دارای

دو تفاوت عمده نسبت به شبکه استفاده ResNet است. تفاوت اول، حذف لایه تماماً متصل موجود در شبکههای ResNet است پارامترهای مربوط به لایههای تماماً متصل تقریباً /۹۰ پارامترهای یک شبکه CNN را شامل می شوند و در نتیجه حجم بسیار زیادی از محاسبات در شبکههای عمیق مربوط به استفاده از این لايهها است. ازاينرو عدم استفاده از اين لايهها سبب کاهش چشمگیر حجم محاسبات و افزایش سرعت پردازش میگردند. از آنجا هدف تخمین مقادیر ارتفاعی برای هر پیکسل در تصویر است، اعمال مرحله کدگشایی بر روی ماتریس ویژگی مناسبتر از اعمال

آن بر روی بردار ویژگی حاصل از لایههای تماماً متصل است.

تفاوت دوم مربوط به اضافه نمودن لایه کانوولوشن به لایههای کمعمق در شبکه ResNet است. عملکرد بهتر در تخمین مقادیر ارتفاعی مربوط به عوارض کوچک ملزوم به داشتن اطلاعات محلى و جزئيات بيشتر است. لایههای کمعمق در شبکههای CNN حاوی اطلاعات محلی و ویژگیهای دارای جزئیات بیشتر میباشند. ازاینرو استخراج ویژگیهای بیشتر در لایههای کمعمق شبکههای CNN میتواند سبب بهبود عملکرد تخمین مقادیر ارتفاعی شود. بهطورکلی تکنیکهای تقویت داده با ایجاد مجموعه دادهای بزرگتر با تولید دادههای جدید و کاربردی بهعنوان یک تنظیم کننده عمل نموده و روند همگرایی را سریعتر مینمایند؛ بهطوریکه سبب جلوگیری از بیشبرازش دادهها شده و امکان عمومی سازی نتایج را بيشتر مىنمايد. اين روش بەصورت كلى شامل اعمال تبدیلاتی بر روی دادههای یا ویژگیهای و یا هر دو میباشد. رایجترین روش افزایش مجازی داده در فضای دادهها صورت گرفته و نمونههای جدید با اعمال تبديلات مختلفى نظير انتقال، چرخش، مقياس گذارى، تغییر فضای رنگی، برش و ... بر روی دادههای موجود بدست میآیند. در تخمین ارتفاع از تکتصویر از روشهای زیر جهت تقویت داده بهره گرفته شده است. دوران: تصاویر ورودی و دادههای ارتفاعی مطابق عددی دلخواه بهصورت تصادفي دوران مي يابند.

انتقال: دادههای ورودی و هدف بهصورت تصادفی بریده شده و به ابعاد ورودی شبکه تغییر ابعاد داده مىشود.

رنگ: باندهای تصویر رنگی ورودی و بهصورت تصادفی در عددی دلخواه ضرب می شوند.

تصویر کردن: دادههای ورودی و هدف بهصورت افقی و عمودی و با احتمال ۰٫۵ تصویر میشوند.

### ۲-۴ شبکه CNN پیشنهادی

از آنجا که هدف رسیدن به تصویر ارتفاعی بوده و هر یک از پیکسلهای تصویر ارتفاعی باید مقداری مجزا داشته باشد نمی توان با شبکههای رایج CNN که درنهایت به لایههای تماماً متصل میرسند به جواب مطلوب رسید. بهعبارتدیگر، شبکههای رایج در کاربردهای شناسایی و کلاسهبندی که تصویر را به تعداد محدودی برچسب تقسیم مینمایند، عملکرد مناسبتی در زمینه مطرح شده ندارند. ازاینرو شبکه CNN پیشنهادی دارای یک ساختار عمیق کدگذار-

<sup>&</sup>lt;sup>2</sup> Encoder-decoder

<sup>&</sup>lt;sup>1</sup> Data augmentation



بهطورکلی، ساختار شبکه در مرحله کدگذاری به این صورت است که ابتدا درصورتیکه تصویر ورودی ۳۲۰۰×۳۲۰۰ باشد، با اعمال یک لایه کانوولوشن و سپس نرمالسازی لایهها لبههای اصلی تشکیلدهنده شکل با جزئیات بالا شناسایی شده و یک ماتریس شکل با جزئیات بالا شناسایی شده و یک ماتریس یک لایه ادغام، سه بلوک باقیمانده نوع دوم و یک بلوک باقیمانده نوع اول ماتریس ویژگی با ابعاد ۲۵۶×۸۰×۸۰

بلوک باقیمانده نوع اول شامل دو شاخه کانوولوشنی است که در شاخه اول آن طی اعمال سه مرحله کانوولوشن به صورت پیاپی تمامی ویژگیهای موجود در تصویر استخراج گشته و در شاخه دوم تنها یک مرحله کانوولوشن به ماتریسها اعمال، نتایج آنها با یکدیگر تلفیق شده خروجی حاصل می گردد. به طورکلی سعی می گردد در بلوک باقیمانده نوع دوم، شاخه اول مشابه بلوک قبلی بوده، ولی در شاخه دوم آن هیچ تابع و لایه ای وجود ندارد. اعمال یک بلوک از نوع دوم به شبکه سبب افزایش عمق شبکه، استخراج خصوصیات بیشتر و حفظ ابعاد ماتریس ویژگی می گردد. در حالی که اعمال بلوک نوع اول همزمان سبب کاهش ابعاد

ماتریس ویژگی نیز میگردد. در شکل (۴) دو بلوک نوع اول و دوم نمایش داده شدهاند.

در مرحله سوم نیز با اعمال سه بلوک از نوع دوم و یک بلوک از نوع اول، ویژگیهای مختلفی استخراج می گردد (۵۱۲×۲۰۰×۲۰۰۹). در مرحله چهارم نیز به همین ترتیب، ضمن کاهش ابعاد ماتریس (۲۰×۲۰) و افزایش عمق فرا (۱۰۲۴)، ویژگیهای کلیتر موجود در سطوح بالاتر و قدرت تفکیک کمتر استخراج میشوند. درنهایت و پس از گذر از مرحله پنجم نقشه ویژگی با ابعاد تمامی لایهها تابع ReLU میشود. لازم به ذکر است که در تمامی لایهها تابع ReLU اعمال شده است؛ چراکه این تمامی لایهها تابع تانژانت هیپربولیک و تصادفی نسبت به توابع تانژانت هیپربولیک و سیگموئید<sup>۲</sup> میشود. به علاوه پیادهسازی بسیار سادهای دارد تنها با یک آستانه گذاری بر روی یک ماتریس صورت می گیرد.

DOI: 10.29252/jgit.7.3.173

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

<sup>&</sup>lt;sup>1</sup> Hyperbolique

<sup>&</sup>lt;sup>2</sup> Sigmoid



شکل ۴: ساختار بلوکهای باقیمانده استفاده شده در مرحله کدگذاری، (الف) بلوک نوع اول، (ب) بلوک نوع دوم

#### ۲-۲-۴ کدگشایی

در مسأله تخمين ارتفاع، بازيابي شكل هندسي عوارض نیز مسالهای مهم و در عین حال دشوار است. چراکه نقشههای ویژگی تولیدشده در طی روند کدگذاری و گذر از چندین مرحله کوچکسازی دارای ابعاد بسیار کوچک می باشند. به طوری که ساختارهای هندسی موجود در منطقه و محل دقیق قرارگیری آنها در تصاویر قابلتشخیص نمی باشد. در شبکه عمیق پیشنهادی ابعاد نقشههای ویژگی پس از گذر از مرحله کدگذاری ۱٬۳۲ ابعاد تصویر ورودی میباشد که این امر نشان از عدم وجود اطلاعات کافی در آنها جهت بازیابی جزئیات دارد. در این مقاله از ساختاری کاملاً كانوولوشن مبنا براى افزايش ابعاد نقشههاى ويژگى استفاده شد [۱۴]. در این ساختار از اتصال نتایج اعمال چهار کانوولوشن متفاوت، تصویری با ابعاد بزرگتر ایجاد می گردد. به طوری که به ترتیب چهار کانوولوشن با ابعاد ۳×۳، ۲×۳، ۳×۲ و ۲×۲ به تصویر اعمال شده و با تعیین دقیق مقادیر لایه گذاری با صفر، چهار ماتریس با ابعادی برابر با ابعاد تصویر ورودی ایجاد می شود. درنهایت با کنار هم قرار دادن این چهار ماتریس، ماتریسی با ابعاد دوبرابر ماتریس تصویر ورودی ایجاد می شود [۱۴]. در این روش به دلیل اعمال عملگرهای

کانوولوشن با ابعاد کوچک و همین طور عدم ایجاد ماتریس های بزرگ در طی روند پردازش که بیشتر پیکسل های آن ها دارای مقدار صفر میباشند، حجم محاسبات در شبکه کاهش یافته و سرعت پردازش افزایش مییابد. نحوه کنار هم قرار دادن نتایج حاصل از اعمال چهار کانوولوشن مطرح شده در شکل (۵) آورده شده است.

همانطور که در ابتدای این بخش ذکر شد، بازیابی هندسه دقیق و ساختار عوارض موجود در منطقه تنها با استفاده از نقشههای ویژگی حاصل از کدگذاری قابل دستیابی نمیباشد. در شبکههای عمیق ویژگیهای سطح پایین و محلی از لایههای کمعمق قابل استخراج هستند [۴۲]. ازاینرو، ساختاری جهت استفاده و بهرهبرداری دوباره از ویژگیهای تولیدشده در مرحله کدگذاری برای بهبود نتایج بزرگ کردن نقشههای ویژگی در مرحله کدگشایی ارائه شده است. در این راستا، ابتدا به هر یک از ماتریسهای نقشه ویژگی موردنظر در مرحله کدگذاری و همینطور ماتریس نقشه ویژگی در کدگشایی که باید اندازه آن افزایش یابد، یک عملگر کانوولوشن با اندازه کرنل ۱×۱ اعمال میشود.

حامد امینی امیرکلائی ، حسین عار فی



شکل ۵: نحوه دو برابر کردن ابعاد نقشههای ویژگی در طی مرحله کدگشایی

اعمال شده بر روی  $\varphi_p \in Q_l$  میباشند.  $\omega_p$  ماتریس وزن روند ارائه شده برای دوبرابر کردن ابعاد نقشههای ویژگی میباشد.  $\oplus$  نشان دهنده عملگر تجمیع و  $\otimes$ نشان دهنده عملگر افزایش ابعاد نقشههای ویژگی میباشند. در شکل (۶) ساختار کلی پیشنهاد شده برای هر مرحله از کدگشایی که شامل افزایش ابعاد نقشههای ویژگی با درنظر گرفتن ویژگیهای تولید شده در مرحله کدگذاری میباشد، نمایش داده شده است. درنهایت روند مطرح شده در شکل(۶) جهت دوبرابر کردن ابعاد نقشههای ویژگی به کارگرفته شد. ساختار ارائه شده برای تلفیق نقشههای ویژگی مراحل کدگذاری و کدگشایی در رابطه(۱) ارائه شده است.  $P_{l+1} = \left( \left( P_l * \omega_p \right) \oplus \left( Q_l * \omega_q \right) \right) \otimes \omega_d$ , (۱) ارائه شده است. کدگذاری و کسی که مراحوانده شده از مرحله کد گذاری و  $P_l$  نقشه ویژگی فراخوانده شده از مرحله کدگذاری و  $P_l$  نقشه ویژگی بدست آمده از مرحله قبل کدگشایی است.  $\omega_p$  و  $\omega_q$  به ترتیب عملگرهای



شکل ۶: ساختار پیشنهاد شده برای کدگشایی

نمایش داده شده است. همان طور که مشخص است، هر چه در مرحله کدگذاری به جلو حرکت میکنیم، ابعاد نقشههای ویژگی کوچکتر شده و صحنه از لحاظ در ادامه برای درک بهتر روند کار در مراحل کدگذاری و کدگشایی، در شکل (۲) از هر یک از مراحل کدگذاری و کدگشایی یک ویژگی یا نقشه ویژگی

با این تفاوت که در مرحله کدگذاری تصاویر دارای ماهیت تصویر رنگی بوده و به نوعی تأثیر گرفته و نمایشدهنده خصوصیات تصویر رنگی ورودی میباشند، درحالی که در مرحله کدگشایی و بهمرور پس از گذر از لایههای طراحی شده، نقشههای ویژگی ساختار ارتفاعی منطقه مورد را نشان میدهند. بصری نامفهومتر می شود. به طوری که لایه های آخر در مرحله کدگذاری تقریباً قابل تفسیر نمی باشند. اما در مرحله کدگشایی معکوس این فرایند صورت می پذیرد. به این ترتیب ابتدا نقشه های ویژگی ابعاد کوچکی داشته و از لحاظ بصری نامفهوم می باشند، اما هر چقدر به انتهای بخش کدگشایی نزدیک می شویم، صحنه واضحتر شده و جزئیات بیشتری به آن اضافه می شود.



شکل ۷: روند کلی اجرای شبکه کدگذار-کدگشا به همراه نمونهای نقشههای ایجادشده در هر مرحله

### ۴–۳– ایجاد *DSM* نهایی

تصاویر ماهوارهای معمولاً مناطق وسیعی با ناهمواریهای زیاد را پوشش می دهند که این سبب می شوند که اتصال تصاویر ارتفاعی تخمین زده شده تنها با استفاده از تصحیح جهش های ارتفاعی در نواحی اتصال کارآمد نباشد. چراکه مدل سازی ناهمواریها در این سطح تنها از طریق آموزش با بهره گیری از این سطح تنها از طریق آموزش با بهره گیری از برش های کوچک امکان پذیر نمی باشد. از این رو، در این بخش روشی جهت استفاده از مدل رقومی ارتفاعی بخش روش ی جهت استفاده از مدل رقومی ارتفاعی ساختار کلی روش پیشنهادی به این صورت است که ابتدا عوارض غیرز مینی از تصاویر ارتفاعی تخمین زده

شده جدا گشته و سپس با افزدون مقادیر ارتفاعی مربوط به عوارض غیرزمینی به مدل رقومی ارتفاعی SRTM، مدل ارتفاعی نهایی استخراج می گردد. الگوریتم پیشنهادی برای استخراج عوراض غیرزمینی از تصاویر ارتفاعی تخمین زده شده در شکل (۸) آورده شده است. در روش پیشنهادی، ابتدا بر اساس محدوده مقادیر تصویر ارتفاعی تخمین زده شده، یک بردار مقادیر تصویر ارتفاعی تخمین زده شده، یک بردار برای شیب در نظر گرفته میشود؛ به این معنا که پیکسلهایی که دارای شیب بیشتری در یک فاصله مشخص نسبت به مقدار تعیین شده باشند، به عنوان پیکسل غیرزمینی در نظر گرفته میشونه



شکل ۸: فلوچارت روش پیشنهادی برای شناسایی عوارض غیرزمینی

شناسایی میشوند. حال سطح تولید شده جایگزین سطح قبلی شده و این روند تا بررسی تمامی مقادیر در بردار جستجو تشکیل داده شده ادامه می یابد. پس از شناسایی عوارض غیرزمینی و حذف پیکسلهای مرتبط با آنها و سپس اعمال درونیابی، DTM مربوط به آن منطقه بدست می آید. سپس با محاسبه تفاضل تصویر ارتفاعی اولیه و DTM بدست آمده عوارض غیرزمینی به همراه مقادير ارتفاعي استخراج مي شوند. درنهايت، پس از کنار هم قرار دادن تصاویر مربوط به عوارض

با ضرب شیب معین شده در بردار مقادیر جستجو، یک حداًستانه ارتفاعی متناسب با هر بردار جستجو تعیین می شود. حال با اعمال عملگر گشایش ٔ مورفولوژی به ازای هر مقدار بردار جستجو و با المان ساختاری دایره شکل به شعاع w یک سطح ایجاد می شود. با محاسبه تفاضل سطح توليد شده از سطح اوليه و اعمال حداًستانه تعیین شده *h*، پیکسلهای غیرزمینی

<sup>1</sup> Dilation

غیرزمینی و اضافه نمودن آن به مدل رقومی ارتفاعی SRTM، مدل ارتفاعی نهایی حاصل میشوند. لازم به ذکر است که از آنجا که ابعاد پیکسل مدل رقومی ارتفاعی SRTM متر میباشد، با استفاده از درونیابی Bilinear ابعاد پیکسل آن با تصویر ماهوارهای مورد مطالعه یکسانسازی میشود.

### ۵- پیادہسازی و ارزیابی نتایج

به منظور پیاده سازی روند پیشنهادی از تصاویر ماهواره ای و DSM مربوط به شهرهای استکهلم (سوئد) و پورتو پرنس (هائیتی) استفاده شده است. مجموعه داده استکهلم شامل یک تصویر ماهواره ای و DSM متناظر آن می باشد که توسط Digital Globe فراهم شده است. تصویر ماهواره ای توسط سنجنده 2 Worldview اخذ شده است که دارای باند

پانکروماتیک با قدرت تفکیک ۴۶/۰ متر و تصویر چندطیفی با قدرت تفکیک ۱/۸۴ متر میباشد. مدل ارتفاعی توسط Vricon تهیه شده است و دارای اندازه پیکسل زمینی برابر با ۵/۰ متر میباشد. مجموعه داده پورتو پرنس شامل یک تصویر ماهوارهای و DSM متناظر آن میباشد. تصویر ماهوارهای مربوط به متر در باند پانکروماتیک و ۲/۴ متر در باند چند طیفی میباشد. DSM منطقه پورتو پرنس توسط سنجنده میباشد. DSM منطقه پورتو پرنس توسط سنجنده یکمتر میباشد. از روش Gram-Schmidt که یکی از روشهای Pan-sharpening که یکی از روشهای اندی با قدرت تفکیک بالا استفاده شد. در شکل (۹) نواحی مطالعاتی نشان داده شدهاند.



(الف) شکل ۹: نواحی مطالعاتی، (الف) تصویر ماهوارهای، (ب) *DSM* 

برای مرحله کدگذاری و همین طور اعمال عملگر معکوس کانوولوشن [۳۹] به جای روند پیشنهادی در مرحله کدگشایی نشان داده شده و با نتایج روند پیشنهادی مقایسه شدهاند.

برای ارزیابی کمی نتایج معیارهای مختلفی شامل، خطای میانگین نسبی (E<sub>R</sub>)[۱۱]، خطای میانگین لگاریتم (E<sub>L</sub>)[۱۱] و ریشه جذر میانگین مربعات (E<sub>RMSE</sub>)[۲۳] ارائه شدهاند که در ادامه روابط آنها آورده شده است (رابطه (۲، ۳ و ۴)).

$$E_{\rm R} = \frac{1}{N} \sum_{i=1}^{n} \frac{|g_i - d_i|}{g_i}, \qquad (1)$$

$$E_{L} = \frac{1}{N}, \sum_{i=1}^{n} \left| \log_{10}^{g_{i}} - \log_{10}^{d_{i}} \right|$$
(٣)

$$E_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{n} (g_i - d_i)^2}, \qquad (f)$$

برای پیادهسازی روش پیشنهادی از یک کامپیوتر با پردازنده Core i7 و کارت گرافیک NVIDIA GeForce GTX 1080 Ti استفاده شد که دارای ۱۱ گیگابات حافظه  $GPU^{1}$  می باشد. مطابق بخش +1، ابتدا فرایند آمادهسازی بر روی تصاویر و دادههای ارتفاعی برای ورودی به شبکه اعمال می شود. به طوری که هریک از تصاویر موردنظر جهت پیادهسازی در ابعاد ۳۲۰×۳۲۰ بريده مى شوند تا مناسب جهت ورود به شبكه عميق باشند. همچنین مقادیر ارتفاعی متناظر با هر تصویر ورودی، از کمینه ارتفاعی خود کم گشته و تصاویر ارتفاعی که حاوی تغییرات ارتفاعی محلی میباشند را حاصل مینماید. درنهایت و پس از تنظیم پارامترهای شبکه عمیق، شبکه آموزش می بیند. درواقع با این کار محدوده تغییرات ارتفاعی محدود گشته و روند همگرایی شبکه تسریع می گردد. از آنجاکه این مجموعه دادهها شامل یک تصویر ماهوارهای و یک مدل ارتفاعی که منطقه وسیع شهری را پوشش میدهد، حدوداً یک سوم (۱/۳) از دادهها به عنوان داده آموزشی و مابقی دادهها برای ارزیابی مورد استفاده قرار گرفتند. تعداد دادههای آموزشی پس از اعمال تبدیلات مربوط به روند تقویت داده به حدود ۴۴ هزار عدد برای ناحیه پورتوپرنس و ۳۸ هزار عدد برای استکهلم میرسد. تعداد ایکها، ابعاد دستهها، روند کاهش وزن، نرخ یادگیری<sup>6</sup> و تکانه<sup>6</sup> برای تمامی شبکهها درهردو ناحیه به ترتیب ۲۰، ۱۶، <sup>۴</sup>-۱۰×۳ و <sup>۳</sup>-۱۰×۵ در نظر گرفته شد. در شکل (۱۰) چند نمونه از تصاویر ارتفاعی تخمین زده شده در این نواحی نشان داده شدهاند. برای ارزیابی بهتر نتایج حاصل از استفاده از شبکههای عمیق دیگر نظیر AlexNet ، [۲۰] و GoogleNet ، [۲۰] و GoogleNet ،

- <sup>1</sup> Graphics Processing Unit
- <sup>2</sup> Epoch
- <sup>3</sup> Batch size
- <sup>4</sup> Weight decay
- <sup>5</sup> Learning rate
- <sup>6</sup> Momentum



شکل ۱۰: نتایج حاصل از تخمین مقادیر ارتفاعی از تک تصویر توسط شبکه پیشنهادی، (الف) تصویر ورودی، (ب) AlexNet (ج). GoogleNet (د) ،VGG، (د) ResNet، (و) شبکه پیشنهادی، (ز) مقادیر ارتفاعی مرجع.

استخراج مدل رقومی سطح با استفاده از تکتصویر ...

حامد امینی امیرکلائی ، حسین عار فی

$T_p$ (m)	$T_t$ (h)	ERMSE (m)	EL	ER	روش	
<i>۰</i> ٬۵۶۲	٩	٧,١٠٧	• ۲۸۱	4,789	Alex Net	استکهلم
٦٫٣٠٣	۲۱	۶٫۷۰۵	۰,۳۶۵	٣,٣٠۴	VGG	
۱٬۰۶۵	١٧	۶,٩۶ <b>۸</b>	۰,۳۵۸	۲,۷۰۰	Google Net	
۱,۷۱۰	74	۶,۱۵۲	•,٣٢٢	۲/۴۳۹	Res Net	
۲۳۴۷	۲۷	۴,۶	•,۲۵۲	۱/۴۵۳	شبکه پیشنهادی	
۰,۵۶۲	٩	٣٫٤٨۵	۰,۲۱۳	• ٫٧٣١	Alex Net	پور تو پرنس پور ا
٣٠٣	۲۵	۳/۵۴۶	۵۱۲٫۰	۰,۶۳۰	VGG	
۱٬۰۶۵	۲.	٣,١٩۵	•,774	۰,۶۴۵	Google Net	
۱,۷۱۰	۳۱	۲٫۸۷۶	۰,۱۹ <b>۸</b>	۰,۵۲۳	Res Net	
۲۳۴۷	۲۷	۲/۱۲۹	•/191	• ، ۳۹ •	شبکه پیشنهادی	

جدول ۱: نتایج ارزیابی حاصل از تخمین مقادیر ارتفاعی

همانطور که در جدول(۱) مشاهده می شود، شبکه پیشنهادی باعث کاهش خطا، بهبود عملکرد و تخمین مقادیر ارتفاعی با دقت بیشتر شده است. همچنین مقایسه کیفی نتایج در شکل(۱۰) نشان از بهبود عملکرد شبکه پیشنهادی در بازسازی بهتر عوارض بهویژه در لبهها و همچنین بازیابی جزئیات بیشتر در هر دو ناحیه مطالعاتی دارد.

همانطور که در بخش ۴-۳ بیان شد، پس از تخمین مقادیر ارتفاعی از تصاویر کوچک بدست آمده، با استفاده از الگوریتم پیشنهادی پیکسلهای زمینی از غیرزمینی جدا میشوند. در این راستا، مقدار شیب و حدآستانه ارتفاعی در الگوریتم پیشنهادی به ترتیب ۳۰ درجه و ۱ متر در نظر گرفته شده اند. پس از شناسایی پیکسلهای غیرزمینی و کنارهم قرار دادن نتایج حاصل شده، یک تصویر یکپارچه که تنها حاوی اطلاعات شده، یک تصویر یکپارچه که تنها حاوی اطلاعات میآید. سپس با اضافه نمودن تصویر بدست آمده به میآید. سپس با اضافه نمودن تصویر بدست آمده به مدل رقومی ارتفاعی *SRTM*، نتیجه نهایی استخراج میشود. در شکل (۱۱ و ۱۲) نتایج نهایی استخراج میشود. در شکل (۱۱ و ۱۲) نتایج نهایی استخراج

نشان داده شدهاند. جهت آنالیز بصری بهتر برخی قسمتهای مدلهای استخراج شده نهایی با بزرگنمایی نمایش داده شدهاند.

سپس DSM نهایی بدست آمده بار دیگر با استفاده از معیار ERMSE ارزیابی شدند که برای ناحیه استکهلم مقدار ۵٬۷۹۶ و برای ناحیه پورتو پرنس مقدار ۳٬۸۵۵ بدست آمد. همان طور که انتظار می رفت، دقت نتایج كاهش يافت كه ناشى از دقت نسبتاً پايين مدل رقومي ارتفاعی SRTM است. چراکه ابعاد پیکسل زمینی در مدلهای رقومی ۳۰ SRTM متر میباشد که عدد بزرگی در مقایسه با خطای الگوریتم پیشنهادی در قطعات كوچك تصويرى (جدول (۱)) مىباشد. بااین حال DSM نهایی بدست آمده چه از لحاظ ساختار کلی که شامل پستی بلندیهای کلی سطح زمین و چه از لحاظ هندسه و ساختار عوارض موجود در مناطق مطالعاتی دارای کیفیت مناسبی میباشند. درواقع به نوعی با ترکیب مدلهای رقومی ارتفاعی SRTM که حاوى اطلاعات بسيار كلى ارتفاعى بوده و اطلاعات ارتفاعي جزئى تر مربوط عوارضي نظير ساختمانها و درختان را شامل نمی شوند.



شکل ۱۱: نتایج حاصل از اتصال تصاویر ارتفاعی و ایجاد یک مدل ارتفاعی پیوسته با استفاده از مدل ارتفاعی SRTM در ناحیه استکهلم، (الف) مدل رقومی ارتفاعی SRTM ، (ب) مقادیر ارتفاعی عوارض غیرزمینی استخراج شده ، (ج) مدل ارتفاعی تخمین زده شده نهایی، (د) مدل ارتفاعی مرجع.

استخراج مدل رقومی سطح با استفاده از تکتصویر ...

حامد امینی امیرکلائی ، حسین عار فی



شکل ۱۲: نتایج حاصل از اتصال تصاویر ارتفاعی و ایجاد یک مدل ارتفاعی پیوسته با استفاده از مدل ارتفاعی SRTM در ناحیه پورتو پرنس، (الف) مدل رقومی ارتفاعی SRTM ، (ب) مقادیر ارتفاعی عوارض غیرزمینی استخراج شده ، (ج) مدل ارتفاعی تخمین زده شده نهایی، (د) مدل ارتفاعی مرجع.

خاکستری ثبت شده می گردد که این عوامل به صورت مستقم بر روی دقت نتایج تاثیر گذارند. عامل دوم که اهمیت بیشتری نیز دارد، تفاوت ساختاری شهرهای پورتوپرنس و استکهلم است. در واقع شهر استکلهم شهری با خانههای غیرمتراکم و پوشش گیاهی زیاد است که در آن خانهها عمدتا دارای سقفهای شیروانی بوده با فاصله از یکدیگر قرار دارند. درحالی که شهر پرتو پرنس، شهری با ترکم بسیار بالا و ساختمانهای به هم پیوسته است که عمدتا دارای سقفهای مسطح می باشد. به طور کلی این دوشهر شباهت ساختاری و هندسی بسیار کمی با یکدیگر دارند و این عامل نیز سبب عملکرد ضعیف در تخمین ساختار کلی ساختمانها می گردد که در شکل (۱۳) نیز مشهود است. برای بهبود نتایج باید کتابخانه آموزشی با استفاده از دادههای مربوط به شهرهایی که دارای شباهت ساختاری با منطقه تست می باشند، تقوت شده و تا حدامکان سعی شود که تصاویر مورد استفاده برای تست مشابه تصاویر مورد استفاده در روند آموزش باشند.





شکل ۱۳: نتایج حاصل از تخمین مقادیر ارتفاعی شهر استکهلم با استفاده از شبکه آموزش دیده توسط دادههای مربوط به شهر پورتوپرنس، (الف) تصویر ورودی، (ب) تصویر ارتفاعی تخمین زده شده، (ج) تصویر ارتفاعی مرجع.

۶- نتیجهگیری

در این مقاله سعی شد تا امکان استخراج اطلاعات سهبعدی و تخمین DSM از تک تصویر ماهوارهای مورد بررسی و آنالیز قرار گیرد. در این راستا، با استفاده از مفاهیم یادگیری عمیق، یک شبکه CNN طراحی شد که پس از آموزش توسط تصاویر ماهوارهای و DSMهای متناظر آنها، توانایی بالایی در زمینه تخمین مقادیر ارتفاعی از تکتصویر داشت. برای استفاده از شبکه پیشنهادی، ابتدا پیشپردازشهای برای آمادهسازی دادههای آموزشی صورت گرفت. شبکه پیشنهادی دارای ساختاری کدگذار -کدگشا می باشد که در مرحله کدگذاری، ویژگیهای متفاوت و قدرتمندی در مقیاسهای متفاوت از تصویر ورودی استخراج شده و در روند کدگشایی به تدریج ویژگیهای استخراج شده با هم تلفیق شده و به مقادیر ارتفاعی تبدیل میشوند. یس از محاسبه مقادیر ارتفاعی برای تصاویر بریده شده، با ارائه یک الگوریتم پیکسلهای زمینی و غیرزمین از هم تفکیک شدند. سیس مقادیر ارتفاعی مربوط به عوارض غیرزمینی شناسایی شده، استخراج و تصاویر حاصل شده به یکدیگر متصل می شوند تا یک سطح ییوسته ایجاد نمایند. در نهایت با افزدون سطح حاصل شده که حاوی مقادیر ارتفاعی مربوط به عوارض غیرزمینی است به مدل رقومی ارتفاعی SRTM با قدرت تفکیک مکانی ۳۰ متر، DSM نهایی منطقه بدست آمد. الگوریتم پیشنهادی با استفاده از تصویر ماهوارهای مربوط به دو شهر استکهلم و پورتو پرنس پیادهسازی و ارزيابي شد. DSM نهايي حاصل شده الگوريتم

#### مراجع

Photogrammetry Remote Sensing and Spatial Information Sciences, vol. 34, pp. 380-385, 2002.

[3] F. Rottensteiner and C. Briese, Automatic generation of building models from LIDAR data and the integration of aerial images: na, 2003.

استخراج مدل رقومی سطح با استفاده از تکتصویر ... حامد امینی امیرکلائی ، حسین عارفی

ییشنهادی بهترتیب مقادیر ۱٬۸۴۹، ۴۱۸ و ۵٬۳۹۶ برای معیارهای E<sub>R</sub>، E<sub>R</sub> و E<sub>RMSE</sub> در ناحیه استکهلم و همینطور به ترتیب مقادیر ۰۰٬۷۶۳ ۲٫۰۵ و ۳٬۸۵۵ رای معیارهای E<sub>L</sub> ،E<sub>R</sub> و E<sub>RMSE</sub> در ناحیه پورتو پرنس حاصل شد. با این که در این تحقیق به دلیل محدود بودن مجموعه داده، تنها از دو تصویر ماهوارهای به همراه مدلهای رقومی دقیق برای پیادهسازی الگوریتم پیشنهادی استفاده شد، نتایج حاصل از تخمین مقادیر ارتفاعی (جدول (۱)) نویدبخش دستیابی به دقتهای بالا در صورت آموزش شبکه پیشنهادی با استفاده از مجموعهای غنی از تصاویر ماهوارهای که پوششهای متنوع سطح زمین را پوشش میدهند، است. همچنین، دسترسی آسان و رایگان بودن مدلهای رقومی ارتفاعی SRTM سبب می شود که عملکرد روند پیشنهادی در استخراج DSM نهایی هیچ محدودیتی نداشته باشد. بهطورکلی، در این تحقیق نشان داده شد که استخراج DSM از تک تصویر ماهوارهای با بهره گیری از توانایی شبکههای عمیق CNN امکانپذیر است. با توجه به تلاش صورت گرفته در این مقاله، همچنان شبکه پیشنهادی در زمینه تخمین ارتفاع عوارض کوچک و مدلسازی تغییرات ارتفاعی کم دچار مشکل است و رفع این مساله نیاز به ارائه روشی جهت استخراج ویژگیهای کارآمدتر به ویژه در لایههای کمعمق است که برای تحقیقات آتی پیشنهاد میشود.

- [1] F. Rottensteiner, "Advanced methods for automated object extraction from LiDAR in urban areas," in Geoscience and Remote Sensing Symposium (IGARSS), IEEE International, pp. 5402-5405, 2012.
- [2] J. Schiewe, "Segmentation of high-resolution remotely sensed data-concepts, applications and problems," International Archives of

- [4] F. Lafarge, X. Descombes, J. Zerubia, and M. Pierrot-Deseilligny, "Structural approach for building reconstruction from a single DSM," IEEE Transactions on pattern analysis and machine intelligence, vol. 32, pp. 135-147, 2010.
- [5] I. V. Florinsky, "Combined analysis of digital terrain models and remotely sensed data in landscape investigations," Progress in Physical Geography, vol. 22, pp. 33-60, 1998.
- [6] H. Murakami, K. Nakagawa, H. Hasegawa, T. Shibata, and E. Iwanami, "Change detection of buildings using an airborne laser scanner," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 54, pp. 148-152, 1999.
- [7] B. P. Olsen, T. Knudsen, and P. Frederiksen, "Digital change detection for map database update," International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences, vol. 34, pp. 357-364, 2002.
- [8] F. H. Sinz, J. Q. Candela, G. H. Bakır, C. E. Rasmussen, and M. O. Franz, "Learning depth from stereo," in Joint Pattern Recognition Symposium, pp. 245-252, 2004.
- [9] J. Skilling and S. Gull, "Algorithms and applications," in Maximum-entropy and Bayesian methods in inverse problems, ed: Springer, pp. 83-132, 1985.
- [10]D. Eigen and R. Fergus, "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture," in Proceedings of the IEEE International Conference on Computer Vision, pp. 2650-2658, 2015.
- [11]D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," Advances in neural information processing systems, pp. 2366-2374, 2014.
- [12]F. Liu, C. Shen, G. Lin, and I. Reid, "Learning depth from single monocular images using deep convolutional neural

fields," IEEE transactions on pattern analysis and machine intelligence, vol. 38, pp. 2024-2039, 2016.

- [13]A. Saxena, S. H. Chung, and A. Y. Ng, "3-d depth reconstruction from a single still image," International journal of computer vision, vol. 76, pp. 53-69, 2008.
- [14]I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab, "Deeper depth prediction with fully convolutional residual networks," 3D Vision (3DV), 2016 Fourth International Conference on, pp. 239-248, 2016.
- [15]Z.-m. Yang and H.-d. Zhao, "A New RBF Reflection Model for Shape from Shading," 3D Research, vol. 8, p. 33, 2017.
- [16]M. A. Rajabi and J. R. Blais, "Improvement of digital terrain model interpolation using SFS techniques with single satellite imagery," International Conference on Computational Science, pp. 164-173, 2002.
- [17]M. A. Rajabi and J. R. Blais, "Optimization of DTM interpolation using SFS with single satellite imagery," The Journal of Supercomputing, vol. 28, pp. 193-213, 2004.
- [18]J. Schennings, "Deep Convolutional Neural Networks for Real-Time Single Frame Monocular Depth Estimation," ed, 2017.
- [19]I. P. Howard, "Perceiving in depth, Vol. 3: Other mechanisms of depth perception," 2012.
- [20]K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
- [21]Z. Zhang, C. Xu, J. Yang, Y. Tai, and L. Chen, "Deep hierarchical guidance and regularization learning for end-to-end depth estimation," Pattern Recognition, vol. 83, pp. 430-442, 2018.
- [22]B. Li, Y. Dai, and M. He, "Monocular Depth Estimation with Hierarchical Fusion of

استخراج مدل رقومی سطح با استفاده از تکتصویر ...

حامد امینی امیرکلائی ، حسین عار فی

Dilated CNNs and Soft-Weighted-Sum Inference," Pattern Recognition, 2018.

- [23]A. Saxena, S. H. Chung, and A. Y. Ng, "Learning depth from single monocular images," Advances in neural information processing systems, pp. 1161-1168, 2006.
- [24]M. Liu, M. Salzmann, and X. He, "Discretecontinuous depth estimation from a single image," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 716-723, 2014.
- [25]S. Srivastava, M. Volpi, and D. Tuia, "Joint height estimation and semantic labeling of monocular aerial images with CNNs," Geoscience and Remote Sensing Symposium (IGARSS), IEEE International, pp. 5173-5176, 2017.
- [26]L. Mou and X. X. Zhu, "IM2HEIGHT: Height estimation from single monocular imagery via fully residual convolutionaldeconvolutional network," arXiv preprint arXiv:1802.10249, 2018.
- [27]P. Ghamisi and N. Yokoya, "IMG2DSM: Height Simulation From Single Imagery Using Conditional Generative Adversarial Net," IEEE Geoscience and Remote Sensing Letters, vol. 15, pp. 794-798, 2018.
- [28]H. A. Amirkolaee and H. Arefi, "Height estimation from single aerial images using a deep convolutional encoder-decoder network," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 149, pp. 50-66, 2019.
- [29]H. A. Amirkolaee and H. Arefi, "Convolutional neural network architecture for digital surface model estimation from single remote sensing image," Journal of Applied Remote Sensing, vol .13, p. 016522, 2019.
- [30]A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," arXiv preprint arXiv:1704.06857, 2017.

- [31]M. Lin, Q. Chen, and S. Yan, "Network in network," arXiv preprint arXiv:1312.4400, 2013.
- [32]Y.-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," Proceedings of the 27th international conference on machine learning (ICML-10), pp. 111-118, 2010.
- [33]A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, pp. 1097-1105, 2012.
- [34]Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436-444, 2015.
- [35]R. H. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung, "Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit," Nature, vol. 405, p. 947, 2000.
- [36]S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv preprint arXiv:1502.03167, 2015.
- [37]S. Wager, S. Wang, and P. S. Liang, "Dropout training as adaptive regularization," Advances in neural information processing systems, pp. 351-359, 2013.
- [38]K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.
- [39]J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431-3440, 2015.
- [40]H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic

نشریہ علمی پژوہشی – مہندسی فناوری اطلاعات مکانی

سال هفتم • شماره سوم • پاییز ۱۳۹۸

segmentation," Proceedings of the IEEE international conference on computer vision, pp. 1520-1528, 2015.

- [41]A. Dosovitskiy, J. T. Springenberg, and T. Brox, "Learning to generate chairs with convolutional neural networks," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1538-154, 2015.
- [42]M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in European conference on computer vision, pp. 818-833, 2014
- [43]C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1-9



Journal of Geospatial Information Technology Vol.7, No.3, Autumn 2019

**Research Paper** 

## Digital surface model extraction with high details using single high resolution satellite image and SRTM global DEM based on deep learning

Hamed Amini Amirkolaee<sup>1</sup>, Hossein Arefi<sup>2\*</sup>

1- PHD student, School of Surveying and Geospatial Eng., College of Eng., University of Tehran, Tehran, Iran 2- School of Surveying and Geospatial Eng., College of Eng., University of Tehran, Tehran, Iran

#### Abstract

The digital surface model (DSM) is an important product in the field of photogrammetry and remote sensing and has variety of applications in this field. Existed techniques require more than one image for DSM extraction and in this paper it is tried to investigate and analyze the probability of DSM extraction from a single satellite image. In this regard, an algorithm based on deep convolutional neural networks (CNN) is designed. In the proposed subject, firstly, some preprocessing such as dividing the satellite image into smaller images, localizing the height values and data augmentation are applied in order to prepare data to enter the network. The proposed CNN network has an encoder-decoder structure in which, different and effective features in different scales are extracted in the encoder stage and the generated features are fused to estimate height values by presenting an effective procedure in the decoding stage. Subsequently, the ground and non-ground pixels are separated and height values of the non-ground objects are extracted. The final DSM is obtained by adding the non-ground pixels with height information to the SRTM digital elevation model (DEM) with 30 meter pixel size. The proposed algorithm is evaluated using the satellite images and their corresponding DSMs. Analyzing the estimated small height images using the proposed CNN indicated 0.921, 0.221 and 2.956m on average for relative mean error  $(E_R)$ , logarithm mean error  $(E_L)$  and root mean squared error (E<sub>RMSE</sub>), respectively. Moreover, analyzing the final seamless DSMs indicated 4.625 on average for Ermse.

Key words: Digital Surface Model, Convolutional Neural Network, single satellite image, SRTM DEM.

Correspondence Address. School of Surveying and Geospatial Eng., College of Eng., University of Tehran, Tehran, Iran Tel: +98 21 61114523, Email: hossein.arefi@ut.ac.ir