

Research Article

## Automatic extraction of urban objects from high-resolution aerial images using convolutional neural networks (Study area: Ahvaz city)

Mostafa Kabolizadeh<sup>1\*</sup>, Kazem Rangzan<sup>1</sup>, Majid Abrehdary<sup>2</sup>, Azim Saberi<sup>1</sup>

1- Department of RS and GIS, Faculty of Earth Sciences, Shahid Chamran University of Ahvaz, Ahvaz, Iran

2- Department of Engineering Science, Division of Mathematics, Computer and Surveying Engineering, University West, Sjöberg, Lars

### 1- Introduction

Due to a large number of structural complications in urban areas, obtaining a model or algorithm for the automatic extraction and reconstruction of these complications from aerial and satellite data can minimize the role of humans in the production of large-scale urban spatial information and significantly reduce the cost and time of their production. Nowadays, the use of UAV technology based on photogrammetric techniques to obtain aerial coverage images with various applications such as environmental and forest monitoring, urban mapping, cultural heritage, mining, and geology has been expanded (Wang, 2021, Onishi, 2018, Granados Bolaños, 2020). Identifying objects from aerial and satellite images using computer image processing techniques and neural networks is one of the most widely used algorithms for identifying urban features (Zheng XU, 2022). Convolution Neural Networks are networked and can extract high-level features in all images.

### 2- Materials and methods

Ahvaz is the capital and the largest city of Khuzestan province. The Storm 2 Vision UAV released the data, which has a Canon M3 camera and was taken with a 24-megapixel resolution. Specialized software (Agisoft Photoscan) was used to prepare orthophotos with appropriate geometrical accuracy and strength from aerial images. In this research, convolutional neural networks are used. These networks are a particular type of neural network composed of convolutional and pooling layers. Artificial neurons are processing units that compute some operation on multiple input variables and typically have an output computed via an activation function. Usually, an artificial neuron has a weight ( $w$ ), which indicates the degree of connection between the artificial neurons, some input variables, and a threshold vector (Watanabe et al., 2020). In this research, a modified linear function has been used as the activation function, called Rectified Linear Unit (ReLU).

The single-shot multibox detector has two main components: a backbone model and an SSD head. The central core is usually a pre-trained image classification network that acts as a feature extractor. Instead of using a moving window, SSD divides the image using a grid, and each grid cell is responsible for detecting objects in that image area. Detecting objects means predicting the class and location of an object in its area. The resNet architecture was first presented in 2015 by He et al. The main idea behind ResNet is that its building blocks are designed to “learn residual functions regarding layer inputs, rather than learning unreferenced functions” (He et al., 2015). In this model, an arbitrary input  $x$  is entered into a neural network layer, and the optimal output  $y$  is considered output.

\* Corresponding author: m.kabolizade@scu.ac.ir

DOI :<https://doi.org/10.2205/AAG.2023.42422.2330>

Received: 2022-12-03

Accepted: 2023-01-02

### 3- Results

The objective of this research is to use a convolutional neural network to automatically identify the spatial boundaries of urban objects from UAV aerial images. Specifically, aerial images of Ahvaz city were analyzed for buildings, roads, and single trees, with dimensions of 256 x 256. The images were selected based on shape, color, texture, and geometric structure and then cropped into squares with a 50% overlap. These segmented images were then used as training data, with at least 250 examples for each class and over 850 examples. Figure 1 displays some training examples taken from the original images.



Fig.1. Samples of training images of the building feature in the study area in two classes

After the model training stage, the proposed trained model has been implemented on a 15-hectare UAV image to extract urban features. Figure 2 shows the final result of applying the model on the studied area. Table 1 also shows the result of evaluating the accuracy of the used model.

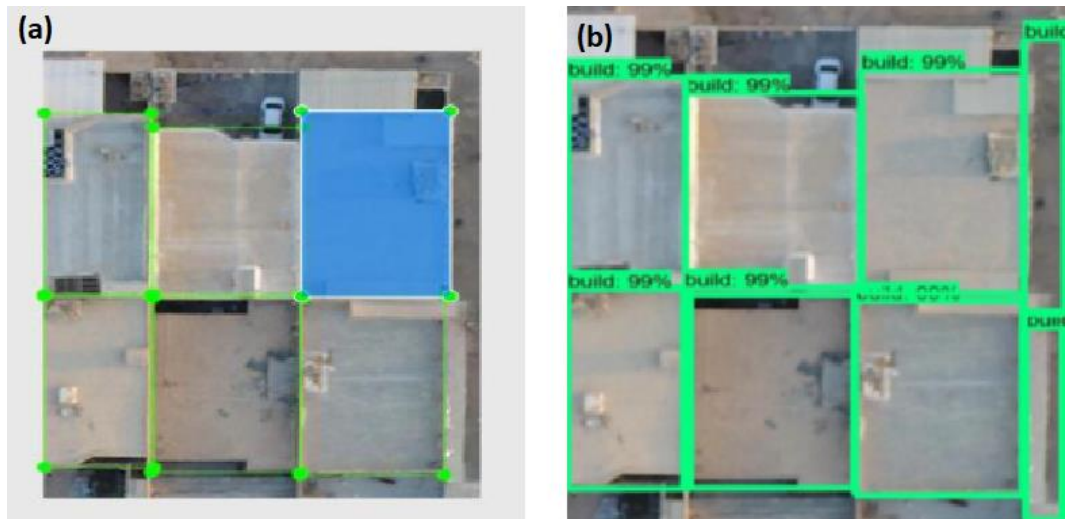


Fig 9. The final result of the model applied to extract building feature in the study area. (a) algorithm output and (b) labeled image.

Table 1. Criteria of model accuracy evaluation

(Precision)	(Recall)	F1 Score
0.86	0.82	0.83

This research uses UAV aerial images with a high spatial resolution to detect and extract urban features automatically. Deep learning methods are among the best methods for recognizing targets from images in today's era. For this purpose, educational examples of building features were taken from aerial images. More than 850 training examples in images with dimensions of 256 x 256 pixels have been prepared for training the model. It has been tried to select the model to be trained with fewer training examples and in less time. The proposed model in this research is a convolutional neural network model based on a single-shot multi-box detector based on the ResNet34 network. 15% of the model's data is also used as validation data. Model training is done with 50 iterations. After training the model, the trained model was implemented on a UAV image within the study area. The images of urban features with a reliability criterion of less than 0.2 have been removed. The confidence range of discovered features is between 20 and 99.87 percent. The evaluation results indicated that the proposed model could extract the desired object with a precision of 0.86, a Recall of 0.82, and an F1Score of 0.83.

## References

- Granados-Bolanos, S., Quesada-Roman, A., Alvarado, G.E., 2020. Low-cost UAV applications in dynamic tropical volcanic landforms. *Journal of Volcanology and Geothermal Research* 410. <https://doi.org/10.1016/j.jvolgeores.2020.107143>
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. *arXiv:1512.03385*. <https://doi.org/10.48550/arXiv.1512.03385>
- Onishi, M., Ise, T., 2018. Automatic classification of trees using UAV onboard camera and deep learning. *ArXiv abs/1804.10390*. <https://doi.org/10.48550/arXiv.1804.10390>
- Wang, X., Xu, N., Wang, B., Liu, Y., Zhang, Sh., 2022. Urban building extraction from high-resolution remote sensing imagery based on multi-scale recurrent conditional generative adversarial network. *GIScience & Remote Sensing* 59, 861-884. <https://doi.org/10.1080/15481603.2022.2076382>
- Watanabe, S., Sumi, K., Ise, T., 2020. Identifying the vegetation type in Google Earth images using a convolutional neural network: a case study for Japanese bamboo forests. *BMC Ecology* 20(65), 1-14. <https://doi.org/10.1186/s12898-020-00331-5>
- Xu, Z., 2022. UAV surveying and mapping information collection method based on Internet of Things. *Internet of Things and Cyber-Physical Systems* 18, 138-144. <https://doi.org/10.1016/j.iotcps.2022.07.002>

### HOW TO CITE THIS ARTICLE:

Kaboli Zadeh, M., Rangzan, K., Saberi, A., 2023. Automatic extraction of urban objects from high-resolution aerial images using convolutional neural networks (Study area: Ahvaz city). *Adv. Appl. Geol.* 13(2), 408-422.

DOI :<https://doi.org/10.2205/AAG.2023.42422.2330>

[https://aag.scu.ac.ir/article\\_18270.html](https://aag.scu.ac.ir/article_18270.html)

©2023 The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, as long as the original authors and source are cited. No permission is required from the authors or the publishers

## استخراج اتوماتیک عوارض شهری از تصاویر با قدرت تفکیک بالای مکانی با استفاده از شبکه‌های عصبی پیچشی (منطقه مورد مطالعه: شهر اهواز)

مصطفی کابلی زاده \*

گروه سنجش از دور و GIS، دانشکده علوم زمین، دانشگاه شهید چمران اهواز

کاظم رنگزن

گروه سنجش از دور و GIS، دانشکده علوم زمین، دانشگاه شهید چمران اهواز

مجید ابره دری

گروه علوم مهندسی، بخش ریاضیات، مهندسی کامپیوتر و نقشه برداری، شوهرگ، لارس

عظیم صابری

گروه سنجش از دور و GIS، دانشکده علوم زمین، دانشگاه شهید چمران اهواز

\* [m.kabolizade@scu.ac.ir](mailto:m.kabolizade@scu.ac.ir)

تاریخ دریافت: ۱۴۰۱/۰۹/۱۲ تاریخ پذیرش: ۱۴۰۱/۱۰/۱۲

### چکیده:

استخراج و بازسازی اتوماتیک عوارض از داده‌های هوایی و ماهواره‌ای می‌تواند نقش انسان را در تولید اطلاعات مکانی بزرگ مقیاس شهری به حداقل رسانده و هزینه و زمان تولید آنها را به شدت کاهش دهد. شناسایی اشیاء هدف از تصاویر هوایی و ماهواره‌ای به خصوص در مناطق شهری با استفاده از تکنیک‌های پردازش تصاویر کامپیوتری و الگوریتم‌های هوش مصنوعی یکی از پرکاربردترین حوزه‌های تحقیقاتی به شمار می‌رود. از انواع این الگوریتم‌ها، شبکه‌های عصبی پیچشی (Convolution Neural Networks) هستند که از توان بالایی برای استخراج ویژگی‌های سطح بالا در انواع تصاویر برخوردارند. در این پژوهش بیش از ۸۵۰ نمونه آموزشی از عوارض ساختاری مهم شهری شامل ساختمان، راه و تک درخت تهیه شده است. در این تحقیق از میان انواع مختلف شبکه‌های موجود، با توجه به سرعت بالاتر انجام پردازش، از شبکه عصبی کانولوشن بر مبنای آشکاساز چندبکس تک شات با پایه شبکه ResNet به منظور کشف و استخراج خودکار محدوده این عوارض استفاده شده است. مدل پیشنهادی توسط نمونه‌های آموزشی بر مبنای ۸۵ درصد داده آموزشی و ۱۵ درصد داده اعتبارسنجی، با ۱۲۰ دوره تکرار و دقت ۹۳ درصد آموزش داده شده است. مدل آموزش دیده بر روی تصاویر مختلف از منطقه مورد مطالعه برای کشف عوارض هدف اجرا شده است. نتایج ارزیابی مدل بر اساس معیار دقت برابر ۰/۸۶ و معیار بازیابی برابر ۰/۸۲ و معیار F1Score برابر ۰/۸۳ می‌باشد. نتایج نشان‌دهنده آنست که الگوریتم پیشنهادی می‌تواند در زمینه‌هایی چون تولید و بروزرسانی نقشه‌های بزرگ مقیاس شهری و همچنین کشف تغییرات در مناطق شهری مورد استفاده قرار گیرد.

**کلمات کلیدی:** شبکه عصبی کانولوشن، آشکاساز چندبکس تک شات، عوارض شهری، ResNet، تصاویر پهپاد

### ۱- مقدمه

در سال‌های اخیر روند افزایش جمعیت شهری و به دنبال آن گسترش و توسعه کالبد شهر، مشکلات زیادی را در تهیه و بروزرسانی اطلاعات مکانی بخصوص در شهرهای بزرگ بوجود آورده است. رشد فزاینده مناطق شهری و دشواری تهیه اطلاعات موثق از این ساختارها در سطح زمین، روش‌های نقشه برداری متداول را به منظور تهیه و بروزرسانی نقشه‌های کاداستر از این مناطق، با چالش جدی مواجه ساخته است. از طرفی امروزه بهبود دقت مسطحاتی نقشه‌ها بسیار مورد توجه قرار گرفته

است.

در حال حاضر، استفاده از تکنولوژی پهپاد مبتنی بر تکنیک‌های فتوگرامتری جهت اخذ تصاویر پوشش‌دار هوایی با کاربردهای مختلف مانند پایش محیط زیست و جنگل‌ها، نقشه برداری شهری، میراث فرهنگی، معدن و زمین شناسی گسترش یافته است (Wang, 2021, Onishi, 2018, Granados-). این امر در نتیجه دسترسی، هزینه‌ی پایین کاربری و اخذ داده‌ها توسط پهپادها می‌باشد. پهپادهای فتوگرامتری با امکان پرواز در ارتفاع پایین و اخذ تصاویر با قدرت

دارد. برای استخراج دقیق مناطق ساختمانی از تصاویر سنجش از دور با وضوح بالا، در این مطالعه، یک روش عملی برای استخراج ساختمان بر اساس شبکه‌های عصبی پیچشی (CNN) پیشنهاد شده است. نتایج تجربی نشان می‌دهد که روش پیشنهادی می‌تواند به طور دقیق و موثر مناطق ساختمان را از تصاویر سنجش از دور استخراج کند. Wang و همکاران (۲۰۲۲) یک مدل بهبود یافته Mask R-CNN برای یادگیری ویژگی‌های پیچیده انواع مختلف سقف ساختمان را پیشنهاد داده‌اند. در این تحقیق ابتدا ادغام شبکه‌های یادگیری عمیق با ویژگی‌های بصری مختلف به منظور تجزیه و تحلیل اثر شاخص پوشش گیاهی (VDVI)، ویژگی‌های تشخیص لبه سوبل و تصاویر مرئی پهپاد بر روی انواع سقف ساختمان‌های روستایی بررسی گردید. در مرحله دوم یک مدل بهبود یافته Mask R-CNN برای یادگیری ویژگی‌های پیچیده استفاده شده است. نتایج نشان می‌دهد مدل مذکور به همراه روش تشخیص لبه Sobel تصاویر RGB بالاترین دقت را دارد. Alsabhan و Alotaiby (۲۰۲۲) در تحقیقی با استخراج عارضه ساختمان بر روی تصاویر ماهواره‌ای با وضوح بالا با استفاده از U-net و ResNet50 پرداخته‌اند. در این راستا شاخص دقت IoU در بخش بندی ساختمان ۸۲٫۲ درصد بوده است. همچنین نتیجه بالایی با F1 score به میزان ۰/۹ به دست آمد. دقت تقسیم بندی تصویر نیز ۹۰٪ برآورد شده است. این مطالعه پتانسیل استخراج خودکار ساختمان با کمک هوش مصنوعی را در مناطق مسکونی با تراکم بالا نشان داد. در نتیجه مشخص گردید که نقشه ساختمان را می‌توان با تصاویر antenna با وضوح بالا و با دقت بالا به دست آورد. Abdollahi و همکاران (۲۰۲۱) در تحقیقی به معرفی دو مدل یادگیری عمیق جدید مبتنی بر خانواده UNet برای تقسیم‌بندی چند عارضه‌ای مانند جاده‌ها و ساختمان‌ها از تصاویر هوایی پرداخته‌اند. محققان بیان کرده‌اند که استخراج ویژگی‌های زمینی، مانند جاده‌ها و ساختمان‌ها از تصاویر هوایی با استفاده از یک سیستم خودکار، کاربردهای زیادی در طیف گسترده‌ای از زمینه‌ها از جمله مدیریت بلایا، تشخیص تغییر، ارزیابی پوشش زمین و برنامه‌ریزی شهری دارد. این کار معمولاً به دلیل شرایط پیچیده مانند مناطق شهری که در آن ساختمان‌ها و اشیاء جاده‌ای توسط سایه‌ها، وسایل نقلیه، درختان و غیره احاطه شده‌اند دشوار است. علاوه بر این، انجام چنین استخراجی توسط متخصصان انسانی به صورت دستی

تفکیک طبیفی و مکانی بالا، در شناخت و شناسایی عوارض شهری بسیار کارآمد هستند. روش‌های فتوگرامتری و سنجش از دور با توجه به وسعت منطقه تحت پوشش از یک طرف و نیز دقت قابل قبول این روش‌ها از طرف دیگر، به عنوان روش‌های مناسب جهت تولید و بهنگام رسانی اطلاعات مکانی و نقشه‌های پوششی شناخته شده‌اند. استخراج و بازسازی عوارض ساخت دست بشر مثل ساختمان‌ها و راه‌ها یکی از زمینه‌های تحقیقاتی فعال در حوزه‌های فتوگرامتری و سنجش از دور و همچنین کامپیوتر و ماشین‌بینایی است. در سال‌های اخیر تلاش‌های بسیاری برای خودکار کردن فرآیند استخراج و بازسازی عوارض شهری از تصاویر رقومی صورت گرفته است، که از آن جمله می‌توان به موارد زیر اشاره نمود. Rakianipour (۲۰۱۹) در پژوهشی که با ایجاد تغییرات در یکی از بهترین راهکارهای اخیر در زمینه تشخیص اشیاء با نام Faster R-CNN، کارایی آن را در تشخیص اشیاء تصاویر هوایی با خصوصیتی مانند اندازه‌ی کوچک اشیاء و تغییرات اندک در ابعاد آن‌ها افزایش داده‌اند. این مدل روی تصاویر هوایی انجام شده و میزان دقت و صحت الگوریتم پیاده‌سازی شده در کلاس‌های مختلف (ساختمان، خیابان و درخت) محاسبه شده است. میزان دقت و صحت در کلاس ساختمان به ترتیب ۰/۹۳ و ۰/۹۷ درصد است و برای کلاس خیابان ۰/۸۷ و ۰/۶۴ درصد و برای کلاس درخت ۱۰۰ و ۰/۷۵ درصد است. Baghani و همکاران (۲۰۱۶) در تحقیق خود امکان استفاده از سکوه‌های بدون سرنشین برای تهیه نقشه‌های کاداستر با مقیاس ۱:۵۰۰ از مناطق شهری وسیع را مورد بررسی قرار دادند. به دلیل وجود نواحی پنهان، اعوجاجات شدیدی در اورتوفتو نهایی در لبه‌ی ساختمان‌ها و دیوارها مشاهده شد. جهت برطرف کردن این مشکل، از نوعی طراحی پرواز ویژه و ابتکاری با عنوان طراحی پرواز مضاعف استفاده شد. نتایج حاصله، افزایش دقت را نشان داد. Wang و همکاران (۲۰۲۲) در مقاله‌ای به مطالعه استخراج ساختمان‌های شهری پرداخته‌اند. چالش بزرگ برای استخراج دقیق ساختمان، تراکم بالا و تفاوت‌های مرزی جزئی در مناطق ساختمان شهری است. اگرچه روش‌های استخراج ساختمان‌های موجود نتایج خوبی در استخراج ساختمان‌های شهری به دست آورده‌اند، اما هنوز مشکلاتی مانند از دست دادن اطلاعات مرزی، استخراج ضعیف برای مناطق متراکم و تداخل جدی سایه‌های ساختمان وجود

کلی (OA) را در مقایسه با سایر تکنیک‌های معمولی فیوژن در سطح ویژگی ارائه می‌دهد.

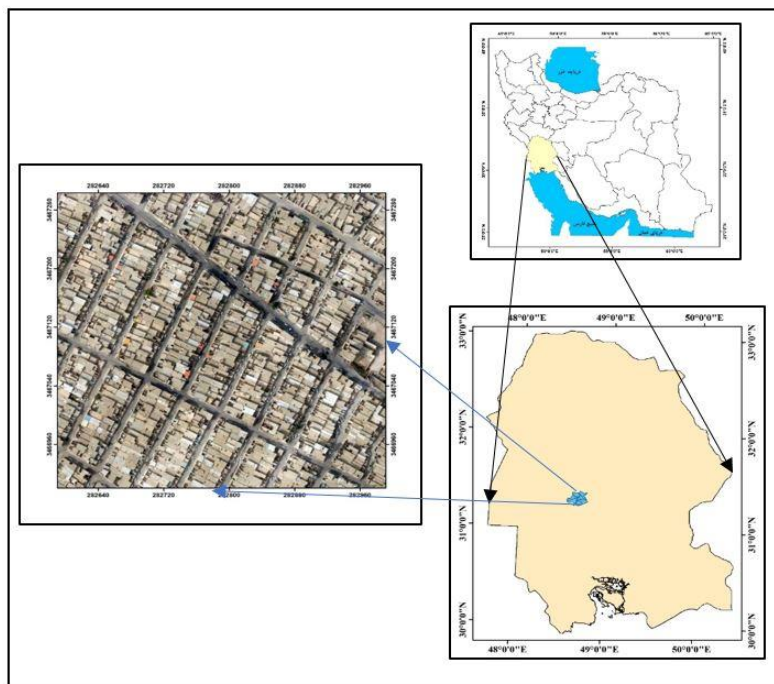
هدف اصلی این تحقیق تشخیص موقعیت عوارض ساختمانی شهری از تصاویر بزرگ مقیاس شهری می‌باشد که در نهایت از آن بتوان نسبت به تهیه و بروزرسانی نقشه‌های شهری اقدام نمود. با دستیابی به این اهداف، داده‌های زیادی برای بروزرسانی این نقشه‌ها بدست خواهد آمد. همچنین بینش جدیدی ارائه می‌دهد که با استفاده از روش‌های مبتنی بر یادگیری عمیق، می‌توان اطلاعات در مقیاس بزرگ را از تصاویر رقومی در دسترس به دست آورد.

## ۲- مواد و روش‌ها

### ۲-۱- منطقه مورد مطالعه و تصاویر هوایی

شهر اهواز یکی از کلان‌شهرهای ایران است، که در بخش مرکزی شهرستان اهواز قرار دارد. جمعیت این شهر طبق سرشماری مرکز آمار ایران در سال ۱۳۹۵ حدود ۱۲۰۰۰۰۰ نفر می‌باشد؛ که به عنوان هشتمین شهر پرجمعیت ایران به‌شمار می‌آید. همچنین اهواز، بزرگترین و پرجمعیت‌ترین شهر جنوب‌غربی کشور است. منطقه مورد مطالعه در موقعیت جغرافیایی ۳۰/۰۰ تا ۳۱/۳۰ درجه شمالی و ۴۸/۲۰ تا ۴۹/۰۰ درجه طول شرقی قرار دارد (شکل ۱).

زمان‌بر و پرهزینه است. مدل‌های ارائه‌شده multi-level bi-directional و context gating UNet (MCG-UNet) و ConvLSTM UNet (BCL-UNet) نامیده می‌شوند. روش‌های پیشنهادی برای تولید نقشه‌های قطعه‌بندی با وضوح بالا و حفظ اطلاعات مرزی حتی در پس‌زمینه‌های پیچیده از مکانیسم‌ها و ویژگی‌های مختلفی استفاده کرده‌اند. مدل MCG-UNet شاخص میانگین دقت F1 را به ترتیب ۱/۸۵٪ و ۱/۱۹٪ و BCL-UNet و UNet در مقایسه با مدل‌های UNet و BCL-UNet برای استخراج دو عارضه جاده و ساختمان بهبود داده است. Daneshtalab و همکاران (۲۰۱۹) از تلفیق تصاویر هوایی و نوری و داده‌های LiDAR و روش شبکه عصبی کانولوشن (CNN) در استخراج توصیف‌کننده‌های سطح بالا (عمیق) برای بهبود طبقه‌بندی استفاده کرده‌اند. در این مقاله، یک چارچوب سطح ویژگی مبتنی بر CNN برای یکپارچه‌سازی داده‌های LiDAR و تصاویر هوایی برای طبقه‌بندی شیء در منطقه شهری پیشنهاد شده است. در این روش، پس از تولید توصیف‌کننده‌های سطح پایین و ادغام آنها در سطح ویژگی توسط لایه بندی، از یک CNN به منظور استخراج ویژگی‌های طیفی و مکانی برای فرآیند طبقه‌بندی استفاده می‌کند، که این عملیات با استفاده از شبکه پرسپترون چند لایه کاملاً متصل انجام می‌شود. نتایج تجربی نشان داد که مدل فیوژن عمیق پیشنهادی حدود ۱۰٪ بهبود در دقت



شکل ۱- منطقه مورد مطالعه

Fig. 1. Study Area

هندسی مناسب از تصاویر هوایی برداشت شده از نرم افزار تخصصی Agisoft Photo Scan استفاده شده است. شکل (۲) اورتوفوتوی تهیه شده و (۳) پهپاد مورد نظر را نشان می دهد.

داده های مورد نظر توسط پهپاد استورم ۲ ویژن که دارای دوربین CanonM3 با رزولوشن ۲۴ مگاپیکسل می باشد، برداشت شده اند. جهت تهیه ی اورتوفوتو با استحکام و دقت



شکل ۲- بخش هایی از اورتوموزاییک تهیه شده از منطقه مورد مطالعه

Fig. 2. some parts of the orthomosaic prepared from the studied area



شکل ۳- پهپاد استورم ۲ ویژن با دوربین CanonM3

Fig. 3. Storm 2 Vision drone with CanonM3 camera

هم پیوسته (نورون های مصنوعی) ارائه می شوند که می توانند مقادیر را از ورودی ها محاسبه کنند و در نتیجه خروجی ای ایجاد می کنند که ممکن است در واحدهای بعدی استفاده شود. نورون های مصنوعی اساساً واحدهای پردازشی هستند که برخی از عملیات را بر روی چندین متغیر ورودی محاسبه می کنند و معمولاً یک خروجی دارند که از طریق تابع فعال سازی محاسبه می شود. به طور معمول، یک نورون مصنوعی دارای وزن  $w_i$  است که نشان دهنده درجه ارتباط بین نورون های مصنوعی، برخی از متغیرهای ورودی  $x_i$  و یک بردار آستانه  $b$  است. از نظر

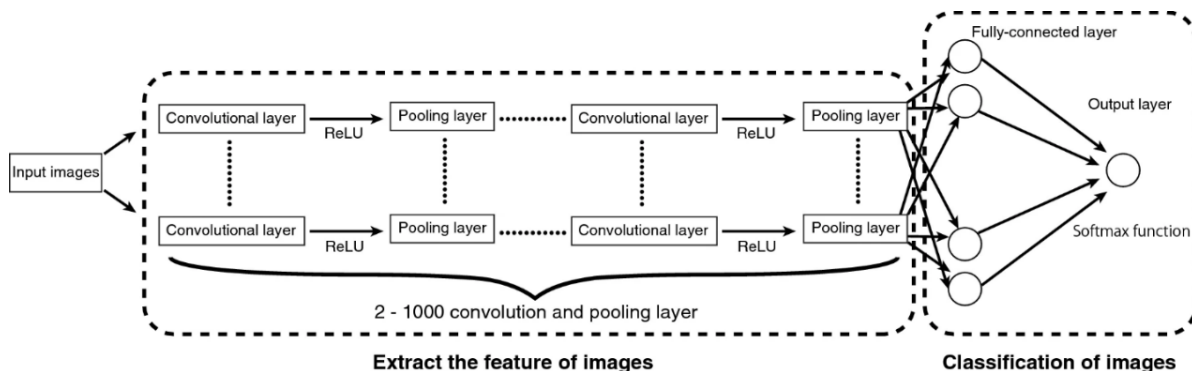
## ۲-۲- شبکه های عصبی پیچشی (CNN)

در این پژوهش، از شبکه های عصبی کانولوشن (شکل ۴) استفاده شده است. این شبکه ها نوع خاصی از شبکه های عصبی پیشخور هستند که از چندین لایه کانولوشنال و لایه های ادغام تشکیل شده اند.

شبکه عصبی پیشخور یک شبکه عصبی مصنوعی است که در آن اتصالات بین گره ها یک چرخه را تشکیل نمی دهند. این شبکه ها که شبیه سازی فعالیت های عصبی در مغز را انجام می دهند، عموماً به عنوان سیستم هایی از واحدهای پردازشی به

صورت معادلات ۱ و ۲ توصیف کرد (Watanabe et al., 2020).

ریاضی، کل ورودی و خروجی نورون‌های مصنوعی را می‌توان به



شکل ۴- نمایش شماییک شبکه‌های کانولوشن (Watanabe et al., 2020)

Fig. 4. Schematic representation of convolution networks (Watanabe et al., 2020)

استخراج می‌کند. بعد از هر لایه کانولوشن، لایه‌های ادغام وجود دارد که برای کاهش واریانس ویژگی‌ها ایجاد می‌شود. معمولاً این بخش با محاسبه عملیات یک ویژگی خاص بر روی یک قسمت از تصویر انجام می‌شود. در نهایت، به دنبال تمام لایه‌های کانولوشن، ادغام و کاملاً متصل، ممکن است از یک لایه طبقه‌بندی کننده برای محاسبه احتمال تعلق کلاس هر تصویر استفاده شود. در این پژوهش از تابع SoftMax استفاده شده است. تابع SoftMax احتمالات هر کلاس هدف را بر روی تمام کلاس‌های هدف ممکن محاسبه می‌کند و به صورت معادله ۴ نوشته می‌شود:

$$y_k = \text{softmax}_k(u_1, u_2, \dots, u_k) = \frac{e^{u_k}}{\sum_{j=1}^k e^{u_j}} \quad (\text{رابطه ۴})$$

که در آن  $k$  تعداد واحدهای خروجی و  $u$  نشان دهنده متغیرهای ورودی است.

برای ارزیابی عملکرد شبکه، باید یک تابع ضرر (loss) نیز تعریف شود. تابع ضرر، اثربخشی شبکه را در مدل‌سازی مجموعه داده آموزشی ارزیابی می‌کند. هدف از آموزش به حداقل رساندن خطای تابع ضرر است. آنتروپی متقاطع تابع SoftMax به صورت معادله ۵ تعریف می‌شود:

$$E = - \sum_{n=1}^N \sum_{k=1}^k t_{nk} \log y_k \quad (\text{رابطه ۵})$$

که  $t$  نشان دهنده بردار داده‌های آموزشی است،  $K$  نشان دهنده کلاس محتمل و  $N$  نشان دهنده تعداد کل نمونه‌ها است.

$$u = \sum_i w_i x_i \quad (\text{رابطه ۱})$$

$$z = f(u + b) = f\left(\sum_i w_i x_i + b\right) \quad (\text{رابطه ۲})$$

که در آن  $u, z, x, w$  و  $b$  به ترتیب کل ورودی، خروجی، متغیرهای ورودی، وزن‌ها و بایاس را نشان می‌دهند. تابع  $f$ ، یک تابع فعال‌سازی را نشان می‌دهد: یک تابع غیرخطی مانند یک تابع خطی، سیگموئید، هذلولی یا یکسو شده.

در این پژوهش از یک تابع خطی اصلاح شده به عنوان تابع فعال‌سازی استفاده شده است و این تابع به عنوان واحد خطی اصلاح شده (ReLU) نامیده می‌شود. تابع ReLU را می‌توان به صورت معادله ۳ تعریف کرد (He et al., 2016):

$$f(u) = \max(0, u) = f \begin{cases} u (u > 0) \\ 0 (u \leq 0) \end{cases} \quad (\text{رابطه ۳})$$

تابع فعال‌سازی ReLU باعث می‌شود خروجی برخی از نورون‌ها صفر شود که باعث پراکندگی شبکه و کاهش وابستگی متقابل پارامترها شده و به طور موثر از مشکل بیش برآزش جلوگیری می‌کند (Xu et al., 2015).

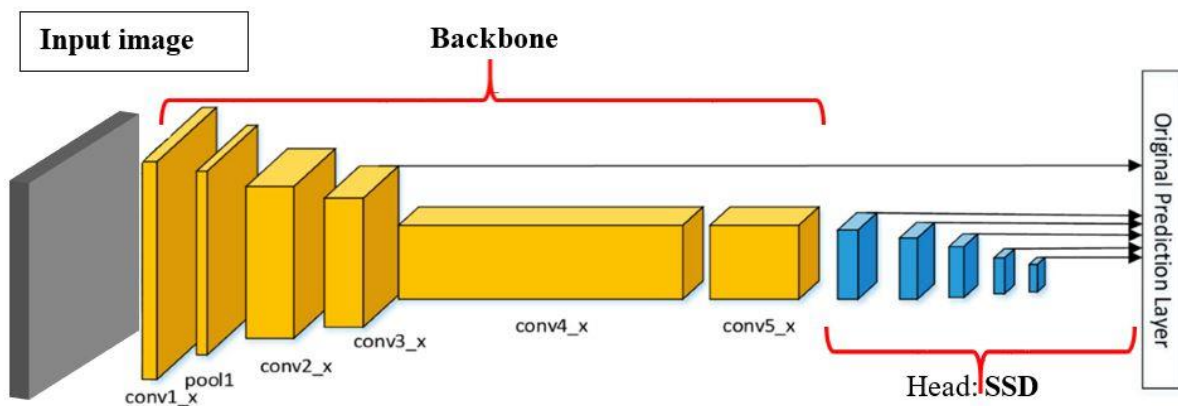
یک شبکه عصبی کانولوشن از یک لایه کانولوشن (convolutional) و یک لایه ادغام (pooling) تشکیل شده است. لایه کانولوشن در استخراج ویژگی‌ها از تصاویر نقش دارد. در این فرآیند، یک فیلتر با اندازه ثابت بر روی تصاویر اعمال می‌شود و الگوهای درجات خاکستری باندها را از تصاویر



### ۳-۲- آشکارساز چندباکس تک شات (SSD)

آشکارساز چندباکس تک شات دارای دو جزء است: یک مدل ستون اصلی (backbone) و سر (SSD head). مدل ستون اصلی معمولاً یک شبکه طبقه‌بندی تصویر از پیش آموزش دیده است که به عنوان استخراج کننده ویژگی عمل می‌کند. SSD به جای استفاده از پنجره متحرک، تصویر را با استفاده از یک شبکه تقسیم می‌کند و هر سلول شبکه‌ای مسئول تشخیص اشیاء در آن ناحیه از تصویر است. تشخیص اشیاء به معنای پیش‌بینی کلاس و مکان یک شیء در آن منطقه است. اگر شیئی وجود نداشته باشد، آن را به عنوان کلاس پس زمینه در نظر گرفته و مکان نادیده گرفته می‌شود. شبکه عصبی عمیق SSD، فضای خروجی جعبه‌های مرزی را به مجموعه‌ای از جعبه‌های پیش‌فرض بر روی نسبت‌ها و مقیاس‌های مختلف در هر مکان

نقشه ویژگی تفکیک می‌کند. در زمان پیش‌بینی، شبکه امتیازهایی را برای حضور هر دسته شیء در هر جعبه پیش‌فرض ایجاد می‌کند و تنظیماتی را برای مطابقت بهتر با شکل شیء ایجاد می‌کند. علاوه بر این، این شبکه پیش‌بینی‌ها را از نقشه‌های چندگانه با وضوح‌های مختلف ترکیب می‌کند تا به طور طبیعی اشیاء با اندازه‌های مختلف را کنترل نماید. این روش نسبت به روش‌هایی دیگر ساده‌تر است زیرا مراحل نمونه‌برداری مجدد پیکسل را حذف می‌کند و تمام محاسبات را در یک شبکه واحد انجام می‌دهد. در نهایت آموزش SSD را آسان می‌کند و ادغام آن در سیستم‌هایی که به یک جزء تشخیص نیاز دارند را راحت می‌کند (Wei et al., 2020). شکل ۵ معماری کلی شبکه عصبی کانولوشن با آشکارساز چندباکس تک شات را نشان می‌دهد. در این پژوهش شبکه طبقه‌بندی تصویر از پیش آموزش دیده، شبکه ResNet می‌باشد.



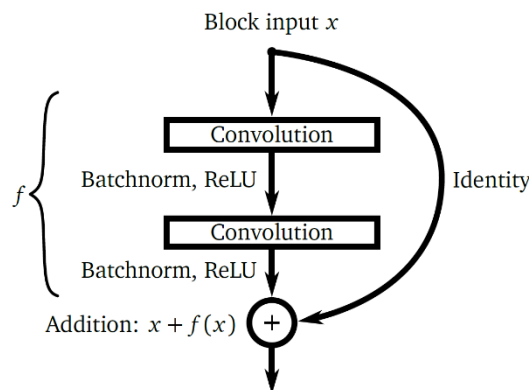
شکل ۵- معماری کلی شبکه عصبی کانولوشن با آشکارساز چندباکس تک شات

Fig. 5. General architecture of convolutional neural network with single-shot multibox detector

### ۴-۲- مدل ResNet

معماری ResNet اولین بار در سال توسط He و همکاران (۲۰۱۵) ارائه شد. ایده اصلی پشت ResNet این است که بلوک‌های ساختمان آن برای «یادگیری توابع باقیمانده با ارجاع به ورودی‌های لایه، به جای یادگیری توابع غیر مرجع» طراحی شده‌اند (He et al., 2015). در این مدل یک ورودی دلخواه  $x$  به لایه‌ای از شبکه عصبی و خروجی بهینه  $y$  در نظر گرفته شده

است. در مورد یک تابع بدون مرجع، شبکه مستقیماً تابع  $f$  را که  $y$  را تولید می‌کند، تقریب می‌زند:  $y = f(x)$ . منظور از تابع باقیمانده در این زمینه، این است که شبکه جدید اکنون  $y = x + f(x)$  را تقریب می‌زند. بنابراین شبکه به جای اینکه یاد بگیرد که چگونه ورودی باید تبدیل شود تا خروجی مورد نظر را ایجاد کند، تفاوت بین ورودی و خروجی بهینه را یاد می‌گیرد. شکل ۶ نشان می‌دهد که چگونه این مفهوم بر روی لایه‌های یک شبکه کانولوشن اعمال می‌شود.



شکل ۶: بلوک ساختمان مدل ResNet (He et al., 2015)

Fig. 6. ResNet model structure blocks (He et al., 2015)

زیاد است. با این حال، اگر مقدار تابع ضرر در طول یادگیری زیاد شود، نشان دهنده بیش برآزش است. معیار دقت (Precision) و معیار بازیابی یا حساسیت (Recall) رایج‌ترین شاخص‌های ارزیابی برای تشخیص اهداف در یادگیری عمیق هستند. هر چه مقدار این معیارها بالاتر باشد، توانایی پیش‌بینی قوی‌تر است. فرمول‌های محاسبه دقت تشخیص و نرخ بازیابی به ترتیب در رابطه ۶ و ۷ نشان داده شده است:

$$P = \frac{TP}{TP+FP} \quad (\text{رابطه ۶})$$

$$R = \frac{TP}{TP+FN} \quad (\text{رابطه ۷})$$

TP، TN و FP بر اساس ماتریس ابهام محاسبه می‌شوند که تعداد تشخیص‌های مختلف را نشان می‌دهند. TP (مثبت واقعی) تعداد اهدافی را نشان می‌دهد که به درستی شناسایی شده‌اند و FP (مثبت کاذب) تعداد اهدافی را نشان می‌دهد که هدف نبوده‌اند، اما به اشتباه به عنوان هدف شناسایی شده‌اند. FN (منفی کاذب) تعداد اهدافی را نشان می‌دهد که هدف هستند، اما شناسایی نشده‌اند. معیار F1-Score یک معادل بین این دو پارامتر ارزیابی مدل است و به عنوان میانگین هارمونیک آنها تعریف می‌شود (معادله ۸):

$$F1_{\text{Score}} = 2 \times \frac{P \times R}{P + R} \quad (\text{رابطه ۸})$$

### ۳- یافته‌ها

همانطور که پیشتر بیان شد، هدف این پژوهش کشف اتوماتیک محدوده مکانی عوارض شهری از تصاویر هوایی پهپاد با استفاده از شبکه عصبی پیچشی می‌باشد. شکل ۷ ساختار مدل

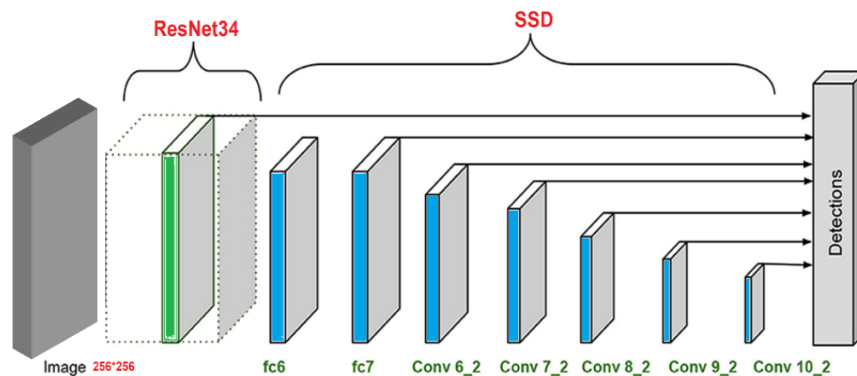
بلوک اصلی مدل ResNet در شکل ۶ نشان داده شده است. بلوک یک ورودی دریافت می‌کند که در آن یک کانولوشن اعمال می‌شود. به دنبال آن نرمال‌سازی دسته‌ای و فعال‌سازی ReLU انجام می‌شود. این کار برای بار دوم تکرار می‌شود و ورودی به نقشه ویژگی حاصل اضافه می‌شود. با این معماری یادگیری تبدیل هویت آسان‌تر است، زیرا بهینه‌ساز فقط باید وزن‌ها را در لایه‌های کانولوشن به صفر برساند (He et al., 2015, 2016). مشاهدات مدل‌های قبلی این بوده که افزودن لایه‌های بیشتر به شبکه موجود لزوماً عملکرد را افزایش نمی‌دهد. نقش کلیدی مدل ResNet ایجاد یک لایه یا بلوک می‌باشد که بتواند به راحتی تغییر هویت را یاد بگیرد. چیدمان این بلوک‌ها در بالای شبکه موجود در حالت ایده‌آل نباید عملکرد آن را بدتر کند، زیرا بلوک‌های جدید تغییر هویت را یاد می‌گیرند و در نتیجه عملکرد مدل پایه را حفظ می‌کنند. اما بلوک‌های جدید اضافه شده تعداد درجات آزادی و عمق شبکه را افزایش می‌دهند. بنابراین ظرفیت یادگیری ویژگی‌های پیچیده را افزایش می‌دهند. در نتیجه اگر نمایش بهتری از داده‌ها وجود داشته باشد، شبکه جدید باید بتواند آن را یاد بگیرد.

### ۵-۲- برآورد دقت یادگیری

مدل در هر دوره یادگیری با استفاده از توابع دقت و ضرر اعتبارسنجی می‌شود. تابع دقت نشان‌دهنده دقت مدل در طبقه‌بندی تصاویر اعتبارسنجی است، در حالی که تابع ضرر نشان‌دهنده عدم دقت پیش‌بینی توسط مدل است. اگر یادگیری مدل موفقیت‌آمیز باشد، مقدار تابع ضرر، کم و مقدار تابع دقت

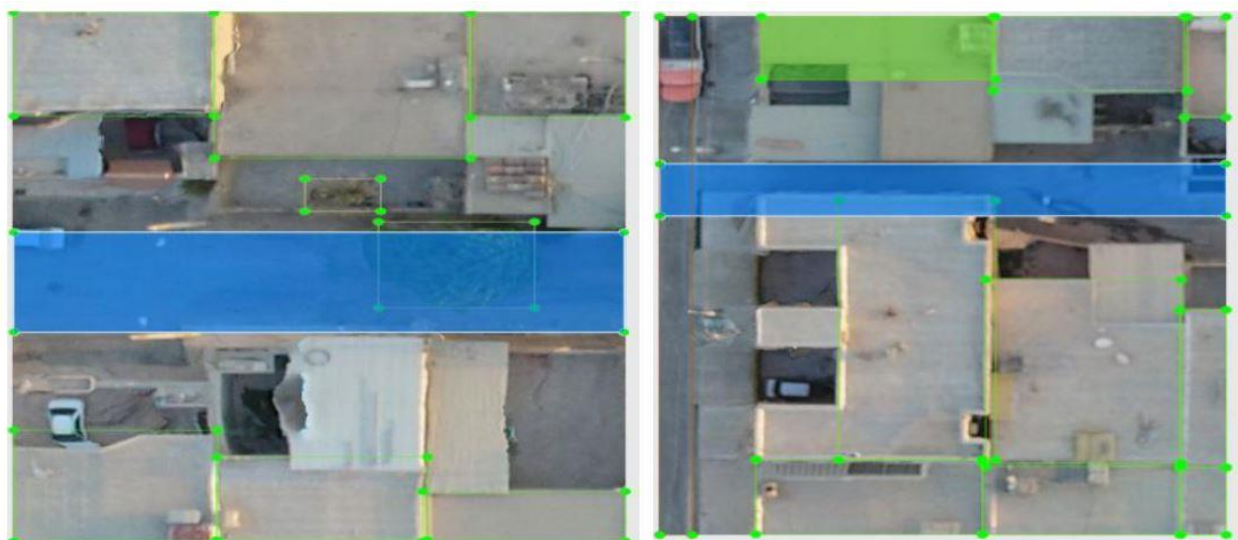
همپوشانی هم به صورت عمودی و هم به صورت افقی بریده شده است. در نهایت تصاویر قطعه قطعه شده به عنوان تصاویر آموزشی استفاده شده است. برای هر کلاس حداقل ۲۵۰ نمونه آموزشی و در مجموع بیش از ۸۵۰ نمونه آموزشی انتخاب شده است. شکل ۸ تعدادی از نمونه‌های آموزشی استخراج شده از تصویر اصلی را نشان می‌دهد.

پیشنهادی را نشان می‌دهد. به این منظور، پس از قطعه‌بندی تصاویر مورد نظر، داده‌های آموزشی آماده می‌گردد (Ise et al., 2018). به بیانی دیگر، تصاویر هوایی برای منطقه مورد مطالعه در شهر اهواز، با ابعاد ۲۵۶ در ۲۵۶ برای سه کلاس ساختمان، راه و تک درخت و بر اساس شکل، رنگ، بافت و ساختار هندسی بریده شده است. در مرحله بعد، این تصاویر به مربع‌های با ۵۰٪



شکل ۷: ساختار مدل پژوهش بر مبنای آشکارساز چندبکس تک شات برای کشف عوارض شهری

Fig. 7 Research model structure based on single-shot multibox detector to detect urban complications



شکل ۸- نمونه‌های تصاویر آموزشی عوارض در منطقه مورد مطالعه

Fig. 8 Examples of training images of objects in the study area

باید برچسب گذاری شوند. عدم قطعیت، سوگیری و خطا در طول فرآیند برچسب‌گذاری بر عملکرد مدل تأثیر می‌گذارد. بدون داده‌های برچسب‌گذاری شده واقعی، قضاوت در مورد

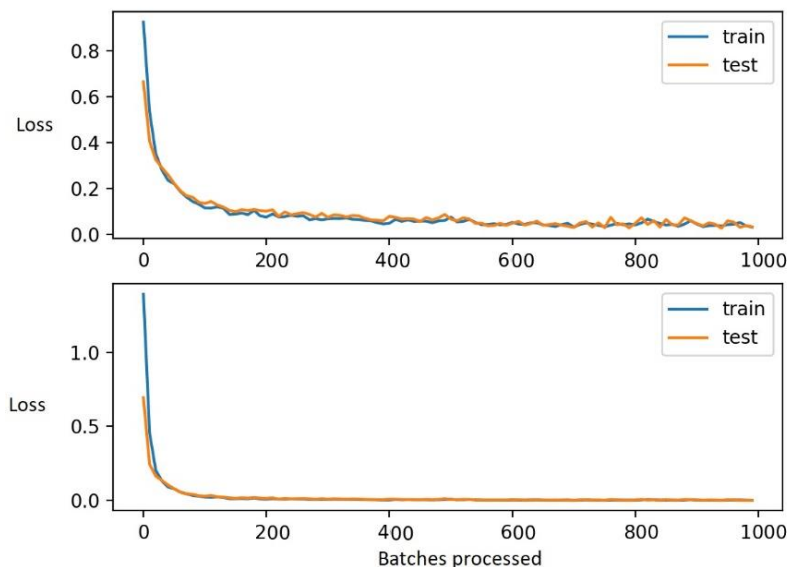
ایجاد مجموعه داده‌های آموزشی بزرگ و با کیفیت بالا یکی از مهمترین مراحل بکارگیری مدل‌های یادگیری عمیق است. به منظور افزایش نرخ فراخوان، اشیاء با سطح اطمینان پایین‌تر

اشیاء در موقعیت مکانی فعال سازی لایه های نهایی تفسیر می-شوند.

پارامترهای مدل اجرا شده در این مطالعه شامل تعداد دوره های آموزشی برابر ۵۰ (epoch=50) و نرخ یادگیری کمتر از ۰/۰۱ می باشد. آموزش مدل پیشنهادی بر روی داده های آموزشی با یک کامپیوتر شخصی لپ تاپ با پردازشگر مرکزی نسل هشت Core i7 و RAM برابر ۱۲ گیگابایت، حدود ۵ ساعت و ۳۶ دقیقه طول کشیده است. میانگین دقت داده های اعتبارسنجی لایه نهایی مدل برابر ۸۹ درصد بوده است. به دنبال افزایش تعداد دوره های آموزشی، مقادیر دقت افزایش و مقادیر تابع ضرر کاهش یافته است. این نتایج نشان می دهد که همه مدل ها بیش از حد به مجموعه داده ها برازش نشده اند و با موفقیت ویژگی های تصاویر قطعه شده را یاد گرفته اند (شکل ۹).

اینکه اشتباهات در داده های آموزشی ممکن است در عملکرد ضعیف مدل نقش داشته باشد، دشوار است. در این پژوهش سعی شده است تعداد نمونه آموزشی بهینه برای دستیابی به دقت بهتر در مدل پیشنهادی انتخاب شود.

برای آموزش از شبکه ResNet 34 استفاده شده است که یک مدل یادگیری عمیق کلاسیک است، به این علت که در این مطالعه از تصاویر با اندازه کوچک به عنوان داده های آموزشی و اعتبارسنجی استفاده می کند. این شبکه با دو لایه کانولوشن، دو لایه ادغام و یک لایه کاملاً متصل ساخته شده است. ResNet34 منجر به تولید ۲۵۶ نقشه ویژگی با ابعاد ۷\*۷ برای یک تصویر ورودی می شود. آشکارساز چندبکس تک شات تنها یک یا چند لایه کانولوشن است که به ResNet34 اضافه شده است و خروجی ها به عنوان جعبه های مرزی و کلاس های

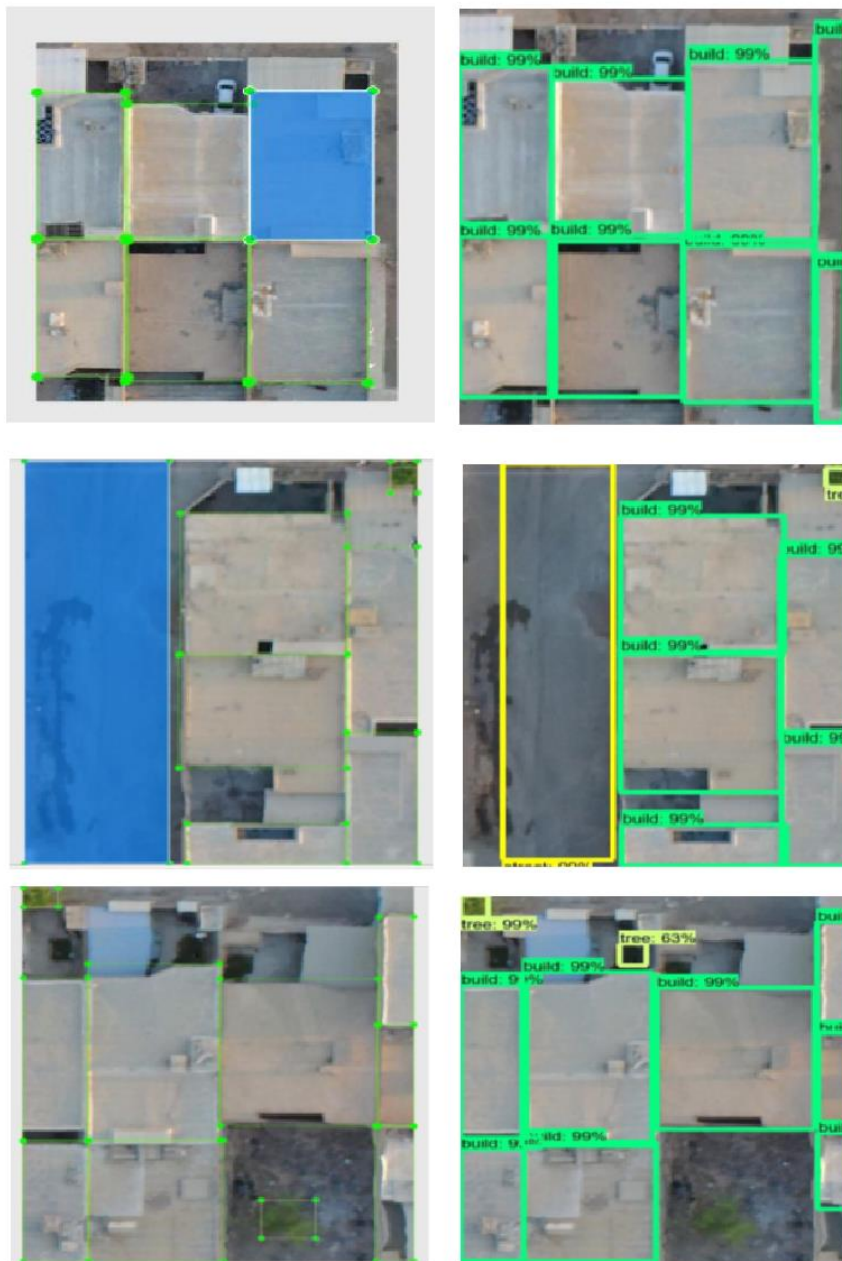


شکل ۹- تابع ضرر بر اساس دوره های یادگیری (دسته های پردازش شده)

Fig. 9 Loss function based on learning periods (processed categories)

دامنه اطمینان عوارض کشف شده بین ۰/۰۳ تا ۲۰/۸۷ درصد می باشد. بنابراین بهترین گزینه های هدف برای اعتبارسنجی الگوریتم تشخیص حفظ شدند. شکل ۱۰ نتیجه نهایی اعمال مدل بر روی منطقه مورد مطالعه را نشان می دهد. جدول ۱ نیز نتیجه ارزیابی دقت مدل استفاده شده را نشان می دهد.

پس از مرحله آموزش مدل، مدل آموزش دیده پیشنهادی بر روی تصویر پهناد اخذ شده به مساحت ۱۵ هکتار برای کشف عوارض شهری اجرا شده است. برای هر عارضه کشف شده توسط مدل، جعبه مستطیلی برای نمایش مکان عوارض بدست آمد. هر جعبه مرزی دارای امتیاز اطمینان از ۰ تا ۱۰۰ است که درصد اطمینان برای حضور عارضه مورد نظر را نشان می دهد.



شکل ۱۰- نتیجه نهایی مدل اعمال شده برای کشف عوارض شهری در منطقه مورد مطالعه. ستون سمت راست) خروجی الگوریتم. ستون سمت چپ) تصویر برچسب گذاری شده

Fig. 10 The result of the applied model to urban objects in the study area. Right) output of the proposed algorithm. Left) labeled image

جدول ۱- معیارهای ارزیابی دقت مدل  
Table 1. Model accuracy evaluation criteria

Precision	Recall	F1 Score
0.86	0.82	0.83

#### ۴- بحث

نقشه عوارض شهری در این پژوهش پیشنهاد می‌شود. در صورت استخراج بصری عوارض از تصاویر نیاز به اپراتور تبدیل و تفسیر تصاویر رقومی وجود دارد و ممکن است اپراتورهای مختلف تشخیص و حتی نحوه استخراج متفاوتی داشته باشند. در این راستا در این پژوهش برای تشخیص و استخراج خودکار عوارض شهری از تصاویر هوایی پهپاد با قدرت تفکیک بالای مکانی استفاده شده است. برای تشخیص اهداف از تصاویر یکی از بهترین روش‌ها در عصر حاضر روش‌های یادگیری عمیق می‌باشد.

در این پژوهش ابتدا نمونه‌های آموزشی عوارض ساختمانی از تصاویر هوایی برداشت شده است. بیش از ۸۵۰ نمونه آموزشی در تصاویر با ابعاد ۲۵۶ در ۲۵۶ پیکسل برای آموزش مدل تهیه شده است. سعی گردیده مدل طوری انتخاب گردد که بتوان آن را با نمونه آموزشی کمتر و در زمان کمتر آموزش داد. مدل پیشنهادی در این پژوهش یک مدل شبکه عصبی کانولوشن بر مبنای آشکارساز چندبکس تک شات بر پایه شبکه ResNet34 می‌باشد. مدل با برنامه‌نویسی پایتون در محیط نرم‌افزار ArcGIS Pro 2.8.6 ایجاد شده است. پس از آن، آموزش مدل بر مبنای داده‌های آموزشی با در نظر گرفتن تصادفی ۸۵ درصد از کل نمونه‌های داده‌ها انجام شده است. ۱۵ درصد داده‌ها نیز به عنوان داده اعتبارسنجی در مدل استفاده شده است. آموزش مدل با ۵۰ تکرار انجام شده است. پس از آموزش مدل، آموزش دیده بر روی یک تصویر پهپاد در محدوده منطقه مورد مطالعه اجرا شده است. تصاویر عوارض شهری که دارای معیار اطمینان کمتر از ۰/۲ داشتند، حذف شده‌اند. دامنه اطمینان عوارض باقیمانده کشف شده بین ۲۰ تا ۹۹/۸۷ درصد می‌باشد. نتایج ارزیابی مدل بر اساس معیار دقت برابر ۰/۸۶ و معیار بازیابی برابر ۰/۸۲ و معیار F1Score برابر ۰/۸۳ می‌باشد.

#### تقدیر و تشکر

بدین وسیله از حمایت مالی معاونت پژوهش و فناوری دانشگاه شهید چمران اهواز در قالب پژوهانه (SCU.EG1401.26151) در انجام این تحقیق تشکر و قدردانی می‌گردد.

همانطور که در شکل ۱۰ دیده می‌شود، الگوریتم پیشنهادی قادر به استخراج محدوده عوارض ساختاری شهری از تصاویر با قدرت تفکیک مکانی بالا در مناطق متراکم شهری است. بسیاری از الگوریتم‌های استخراج اتوماتیک عوارض، در مناطق نسبتاً کم‌تراکم شهری مورد پیاده‌سازی و ارزیابی قرار گرفته‌اند؛ در حالیکه در تحقیق حاضر یکی از مناطق شهری با بافت متراکم و تنوع مناسب جهت ارزیابی مورد استفاده قرار گرفته است. یکی از مهمترین مزایای الگوریتم استفاده شده در این تحقیق، وابستگی اندک آن به پارامترهایی است که توسط کاربر تعیین می‌گردد. این امر ضمن افزایش میزان خودکار بودن فرآیند استخراج، موجب کاهش خطای انسانی در این فرآیند خواهد شد. از سوی دیگر، لازم به ذکر است که در روش پیشنهادی، محدوده عوارض استخراج شده است و در برخی عوارض، به دلیل پیچیدگی شکل، برخی قسمت‌ها خارج از محدوده استخراج شده قرار دارند و قسمت‌هایی هم به اشتباه بیشتر نشان داده است. این امر که شاید بتوان آنرا به عنوان یکی از محدودیت‌های مدل‌های شبکه عصبی پیچشی در نظر گرفت، موجب می‌گردد عملکرد الگوریتم در برخی مناطق نیاز به پس‌پردازش و تصحیح داشته باشد.

#### ۵- نتیجه‌گیری

همانطور که پیشتر نیز بیان گردید، استخراج اتوماتیک موقعیت مکانی ساختمان‌ها ضمن در اختیار قرار دادن اطلاعاتی مفید جهت استفاده در کاربردهای برنامه‌ریزی، مدیریت و غیره، به عنوان مقدمه مرحله بازسازی ساختمان نیز به شمار می‌رود. روش‌های زمینی تهیه نقشه ساختمان‌ها با توجه به شرایط آب و هوایی و نیاز به صرف هزینه و زمان بالا، هر چند دارای دقت بالایی می‌باشند، مقرون به صرفه نیستند. لذا راه‌کار جایگزین استفاده از تصاویر هوایی (مانند: پهپاد) یا تصاویر ماهواره‌ای با قدرت تفکیک بالا است. در استفاده از تصاویر ماهواره‌ای و هوایی کلاسیک علاوه بر هزینه تهیه تصاویر، شامل محدودیت‌های زمانی و مکانی خواهد بود. در این راستا استفاده از تصاویر پهپاد به دلیل در دسترس بودن، برداشت نسبتاً ارزان قیمت و همچنین قدرت تفکیک مکانی کاملاً مناسب برای تشخیص و استخراج

مراجع

- Abdollahi, A., Pradhan, B., Shukla, N., Chakraborty, S., Alamri, A., 2021. Multi-Object Segmentation in Complex Urban Scenes from High-Resolution Remote Sensing Data. *Journal of Remote Sensing* 13(18), 1-22. <https://doi.org/10.3390/rs13183710>
- Alsabhan, W., Alotaiby, T., 2022. Automatic Building Extraction on Satellite Images Using Unet and ResNet50. *Hindawi, Computational Intelligence and Neuroscience* 2022, 1-12. <https://doi.org/10.1155/2022/5008854>
- Baghani, A., Valadan-zoej, M.J., Mokhtarzadeh, M., 2016. Large-scale cadastral map production from urban areas using UAV photogrammetry. *The 1st National Conference on Geospatial Information Technology*. 1-17. <https://civilica.com/doc/461639>
- Daneshtalab, S., Rastiveis, H., Hosseiny, B., 2019. CNN-Based Feature-Level Fusion of Very High Resolution Aerial Imagery and LIDAR Data. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 279-284.
- Granados-Bolanos, S., Quesada-Roman, A., Alvarado, G.E., 2020. Low-cost UAV applications in dynamic tropical volcanic landforms. *Journal of Volcanology and Geothermal Research* 410. <https://doi.org/10.1016/j.jvolgeores.2020.107143>
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. *arXiv:1512.03385*. <https://doi.org/10.48550/arXiv.1512.03385>
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- Onishi, M., Ise, T., 2018. Automatic classification of trees using UAV onboard camera and deep learning. *ArXiv abs/1804.10390*. <https://doi.org/10.48550/arXiv.1804.10390>
- Rakianipour, S., 2019. Automatic building extraction from digital images using artificial intelligence method. MSc thesis. Shahid Chamran University of Ahvaz.
- Wang, X., Xu, N., Wang, B., Liu, Y., Zhang, Sh., 2022. Urban building extraction from high-resolution remote sensing imagery based on multi-scale recurrent conditional generative adversarial network. *GIScience and Remote Sensing* 59, 861-884. <https://doi.org/10.1080/15481603.2022.2076382>
- Wang, Y., Li, Sh., Teng, F., Lin, Y., Wang, M., Cai, H., 2022. Improved Mask R-CNN for Rural Building Roof Type Recognition from UAV High-Resolution Images: A Case Study in Hunan Province, China. *Journal of Remote Sensing* 14(2), 1-25. <https://doi.org/10.3390/rs14020265>
- Wang, K., Meng, Z., Wu, Z., 2021. Deep Learning-Based Ground Target Detection and Tracking for Aerial Photography from UAVs. *Applied Sciences* 11(18), 1-11. <https://doi.org/10.3390/app11188434>
- Watanabe, S., Sumi, K., Ise, T., 2020. Identifying the vegetation type in Google Earth images using a convolutional neural network: a case study for Japanese bamboo forests. *BMC Ecology* 20(65), 1-14. <https://doi.org/10.1186/s12898-020-00331-5>
- Xu, B., Wang, N., Chen, T., Li, M., 2015. Empirical evaluation of rectified activations in convolutional network. *ArXiv Prepr. arXiv:1505.00853*. <https://doi.org/10.48550/arXiv.1505.00853>
- Xu, Z., 2022. UAV surveying and mapping information collection method based on Internet of Things. *Internet of Things and Cyber-Physical Systems*, 18, 138-144. <https://doi.org/10.1016/j.iotcps.2022.07.002>