



قطعه بندی معنایی تصاویر پهپاد- مبنای نواحی شهری بر مبنای

شبکه عصبی رمزگذار- رمزگشا U-Net



ارائه دهنده:

عباس مجیدی زاده

اداره پردازش تصاویر و سنجش از دور

تابستان ۱۴۰۳



SEMANTIC SEGMENTATION OF UAV IMAGES BASED ON U-NET IN URBAN AREA

A. Majidizadeh ^{1,*}, H. Hasani ¹, M. Jafari ¹

¹ Department of Geodesy and Surveying Engineering, Tafresh University, 79611-39518, Tafresh, Iran – (h.hasani, Jafari, a.majidizadeh)@tafreshu.ac.ir

Commission IV, WG IV/3

KEY WORDS: Semantic Segmentation, UAV, Deep Learning, Convolutional Neural Network, Encoder-Decoder Architecture, U-Net.

ABSTRACT:

Semantic segmentation of aerial data has been one of the leading researches in the field of photogrammetry, remote sensing, and computer vision in recent years. Many applications, including airborne mapping of urban scenes, object positioning in aerial images, automatic extraction of buildings from remote sensing or high-resolution aerial images, *etc.*, require accurate and efficient segmentation algorithms. According to the high potential of deep learning algorithms in the classification of complex scenes, this paper aims to train a deep learning model to evaluate the semantic segmentation accuracy of UAV-based images in urban areas. The proposed method implements a deep learning framework based on the U-Net encoder-decoder architecture, which extracts and classifies features through layers of convolution, max pooling, activation, and concatenation in an end-to-end process. The obtained results compare with two traditional machine learning models, Random Forest (RF) and Multi-Layer Perceptron (MLP). They rely on two steps that involve extracting features and classifying images. In this study, the experiments are performed on the UAVid2020 semantic segmentation dataset from the ISPRS database. Results show the effectiveness of the proposed deep learning framework, so that the U-Net architecture achieved the best results with 75.15% overall accuracy, compared to RF and MLP algorithms with 52.51% and 54.65% overall accuracy, respectively.

☐ قطعه‌بندی معنایی؟

☐ کاربردهای قطعه‌بندی معنایی در فتوگرامتری و سنجش از دور؟

☐ هدف از قطعه‌بندی معنایی تصاویر هوایی شهری؟

☐ عملکرد روش‌های سنتی یادگیری ماشین در مقایسه با رویکردهای یادگیری عمیق؟

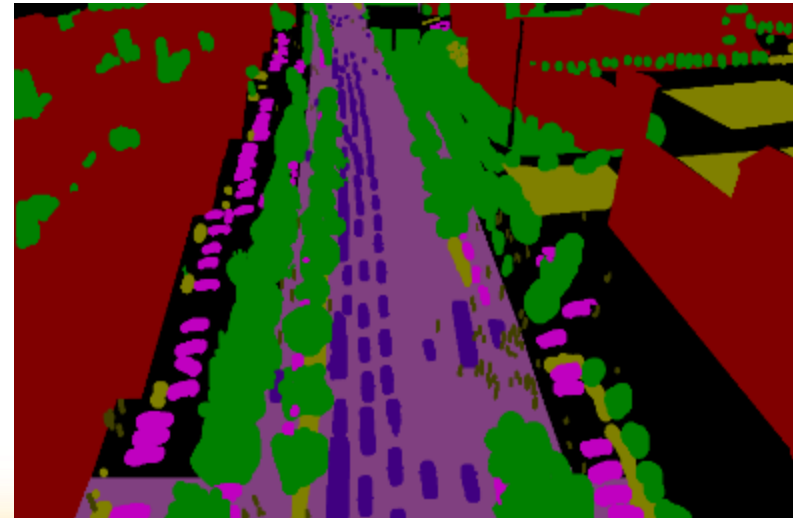
مقدمه

بیان مسئله و
ضرورت تحقیق

روش‌شناسی
تحقیق

پیاده‌سازی‌ها و
تحلیل نتایج

نتیجه‌گیری و
پیشنهادها



قطعه‌بندی معنایی تصاویر پهپاد-مناوهای شهری بر مبنای شبکه عصبی رمزگذار-رمزگشا U-Net



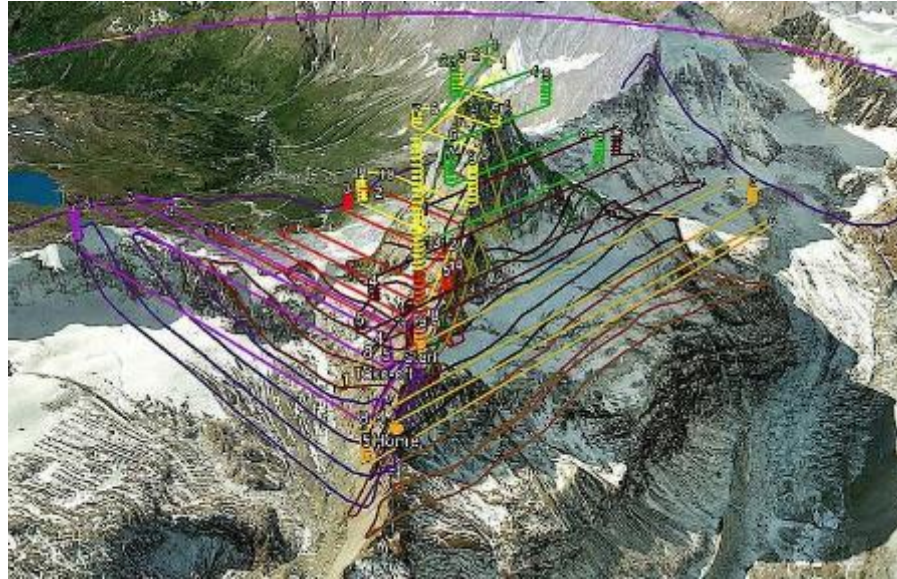
مقدمه

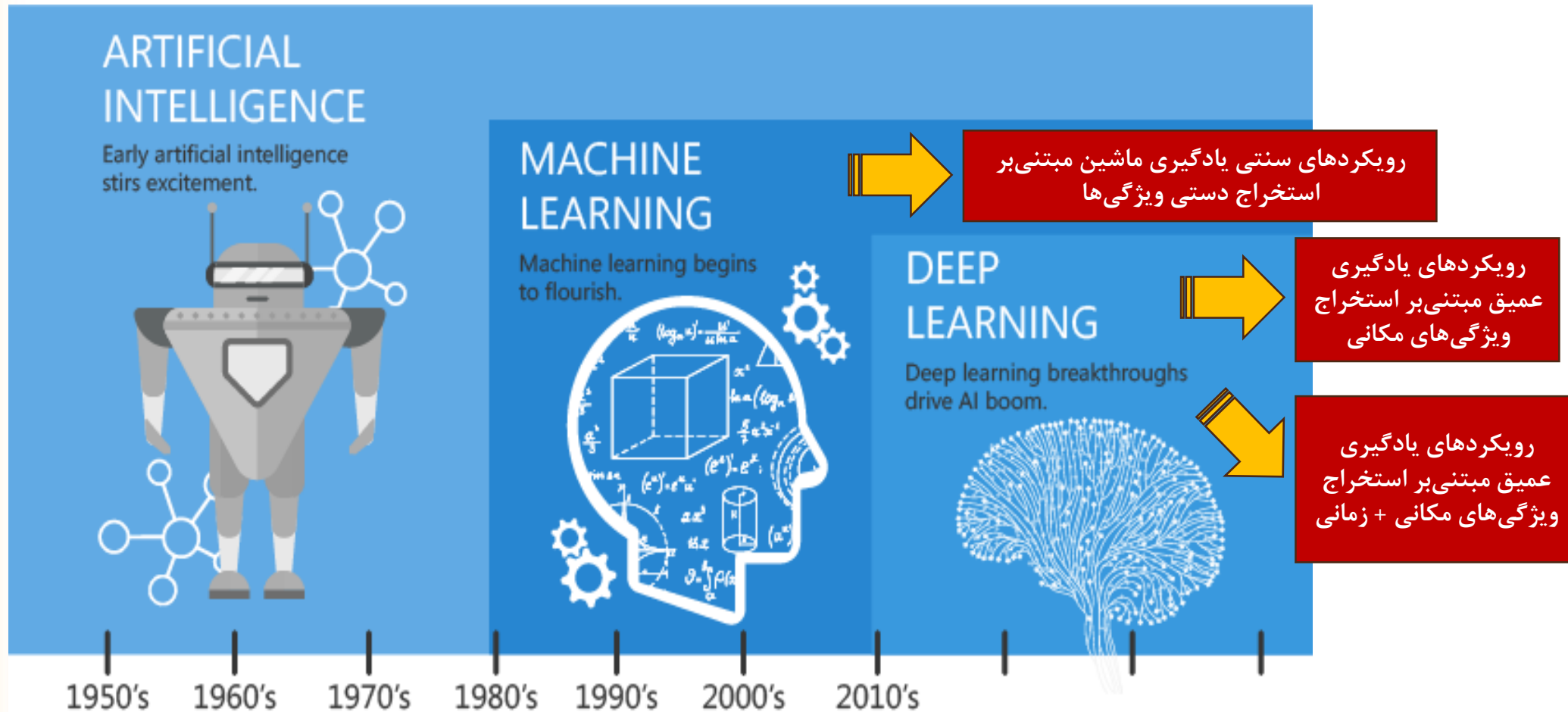
بیان مسئله و
ضرورت تحقیق

روش‌شناسی
تحقیق

پیاده‌سازی‌ها و
تحلیل نتایج

نتیجه‌گیری و
پیشنهادها





مقدمه

بیان مسئله و
ضرورت تحقیق

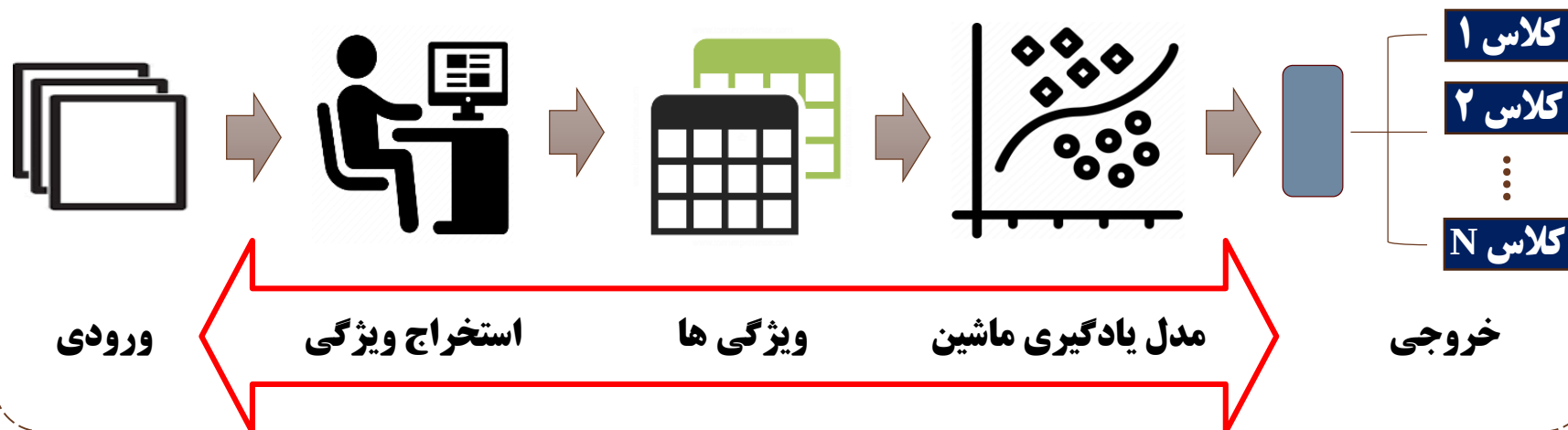
روش‌شناسی
تحقیق

پیاده‌سازی‌ها و
تحلیل نتایج

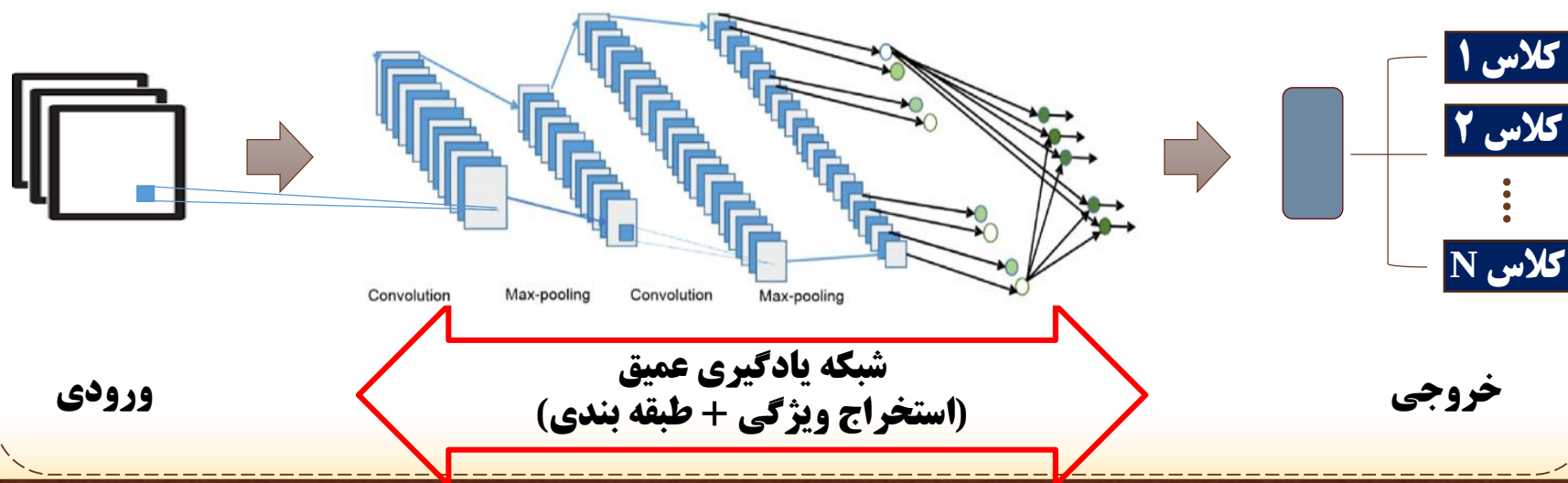
نتیجه‌گیری و
پیشنهادها

قطعه بندی معنایی تصاویر پهلو-مناواحی شهری بر مبنای شبکه عصبی رمزگذار-رمزگشا U-Net

روش سنتی یادگیری ماشین



روش یادگیری عمیق



یادگیری ماشین

مقدمه

بیان مسئله و
ضرورت تحقیق

روش شناسی
تحقیق

پیاده سازی ها و
تحلیل نتایج

نتیجه گیری و
پیشنهادها

تعریف مسئله

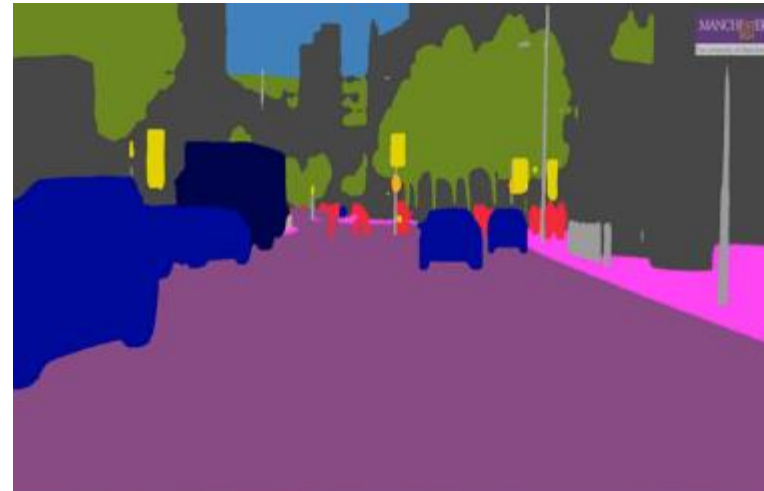
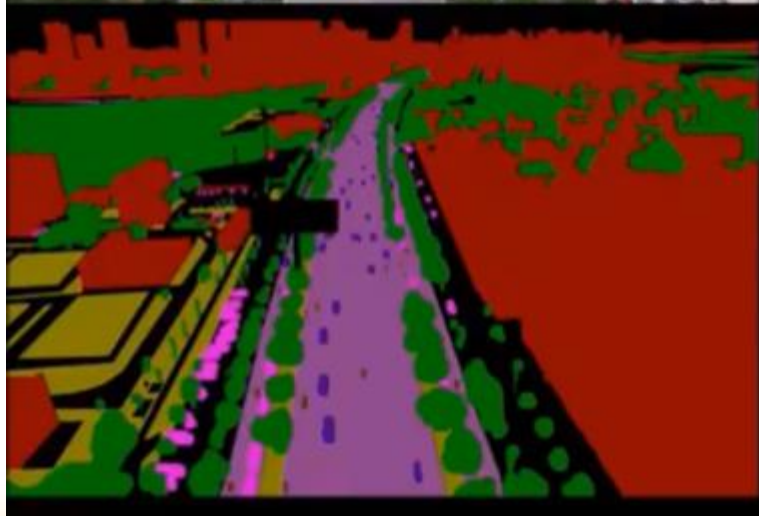
مقدمه

بیان مسئله و
ضرورت تحقیق

روش‌شناسی
تحقیق

پیاده‌سازی‌ها و
تحلیل نتایج

نتیجه‌گیری و
پیشنهادها



ضرورت تحقیق

مقیاس
کوچک

مقیاس
بزرگ

6/23



مقدمه

بیان مسئله و
ضرورت تحقیق

روش‌شناسی
تحقیق

پیاده‌سازی‌ها و
تحلیل نتایج

نتیجه‌گیری و
پیشنهادها

اهداف تحقیق

- 1) استخراج ویژگی‌های مکانی فریم‌های ویدئویی بر مبنای معماری یادگیری عمیق کانولوشنی برای قطعه‌بندی معنایی تصاویر هوایی شهری.
- 2) تجزیه و تحلیل عملکرد قطعه‌بندی معنایی فریم‌های ویدئویی معماری کانولوشنی در مقایسه با روش‌های معمول یادگیری ماشین.

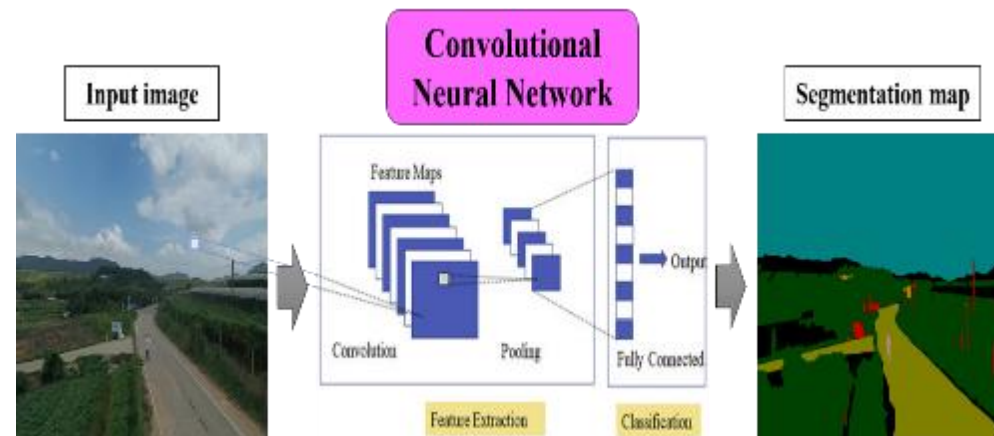
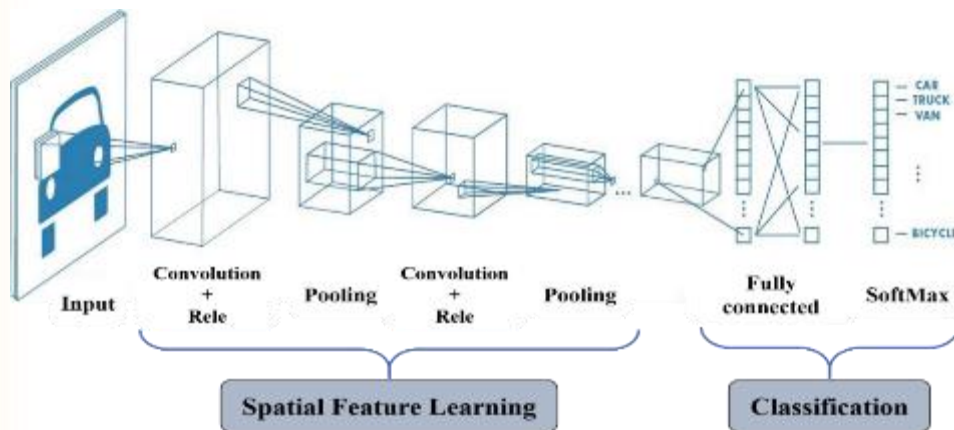
مقدمه

بیان مسئله و
ضرورت تحقیق

روش‌شناسی
تحقیق

پیاده‌سازی‌ها و
تحلیل نتایج

نتیجه‌گیری و
پیشنهادها





سوالات تحقیق

مقدمه

بیان مسئله و
ضرورت تحقیق

روش‌شناسی
تحقیق

پیاده‌سازی‌ها و
تحلیل نتایج

نتیجه‌گیری و
پیشنهادها

1. آیا معماری رمزگذار-رمزگشا کانولوشنی عملکرد مناسبی در قطعه‌بندی معنایی پهپاد-مبنا نواحی شهری ارائه می‌دهد؟
2. عملکرد مدل یادگیری عمیق در مقایسه با رویکردهای سنتی قطعه‌بندی معنایی چگونه است؟



طبقه بندی معنایی تصاویر پهپاد-مناوهای شهری بر مبنای شبکه عصبی رمزگذار-رمزگشا U-Net

فلوچارت روش پیشنهادی

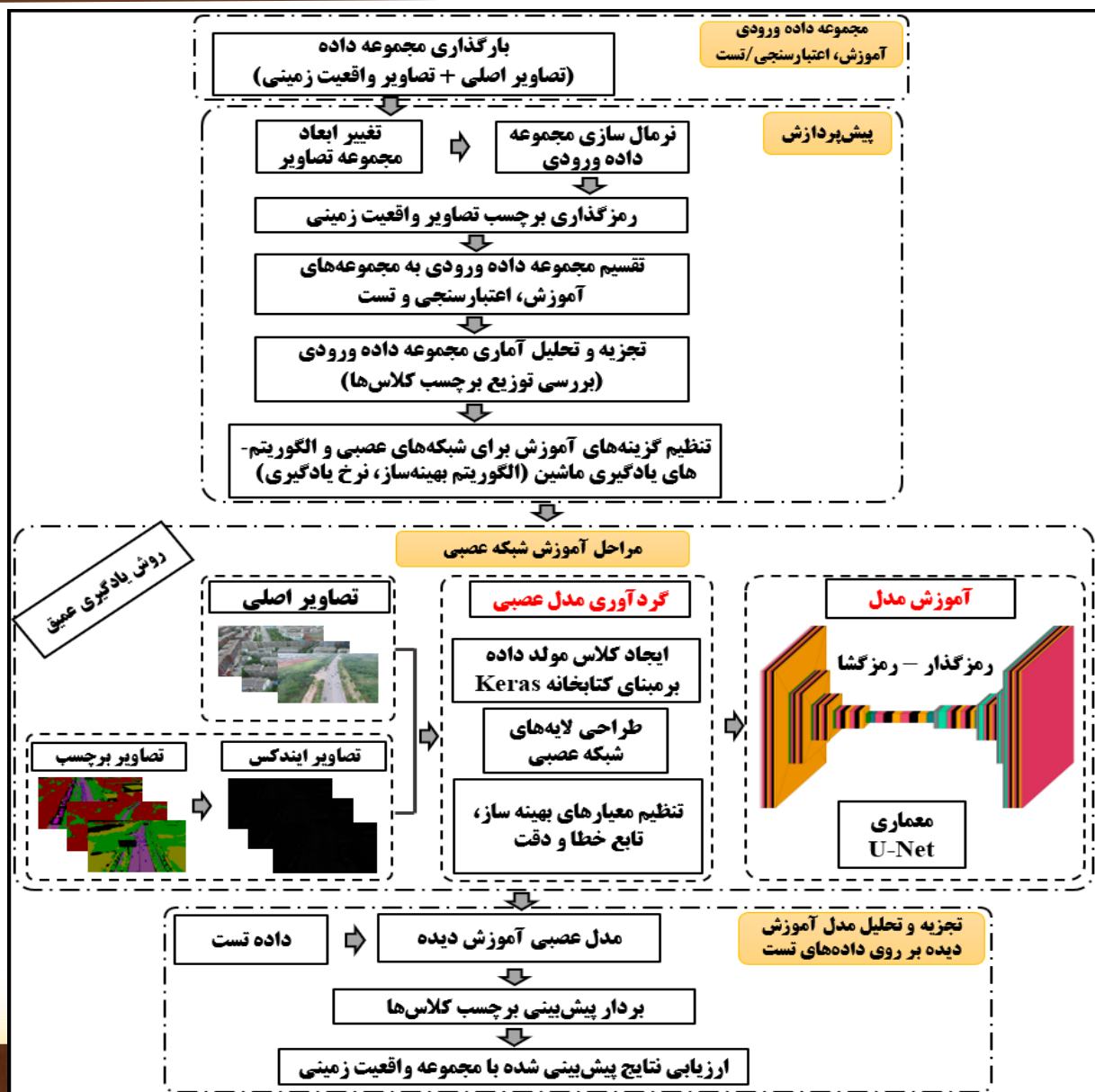
مقدمه

بیان مسئله و
ضرورت تحقیق

روش شناسی
تحقیق

پیاده سازی ها و
تحلیل نتایج

نتیجه گیری و
پیشنهادها



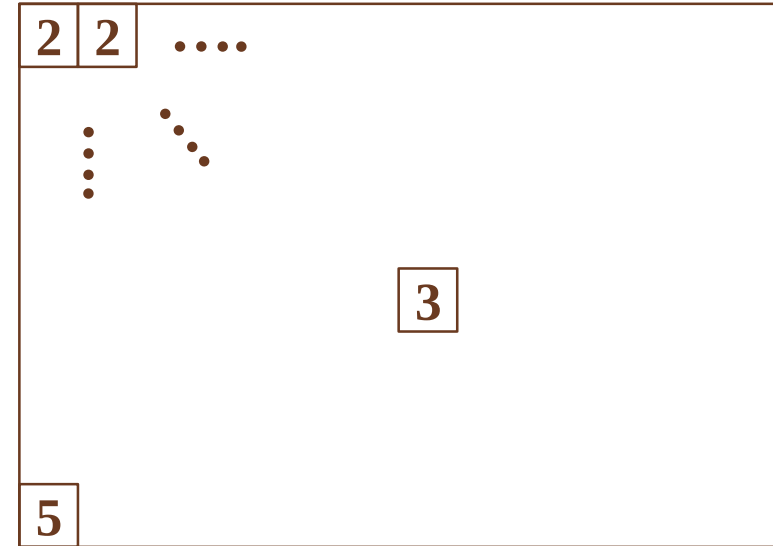
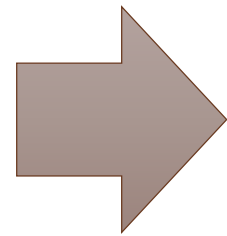
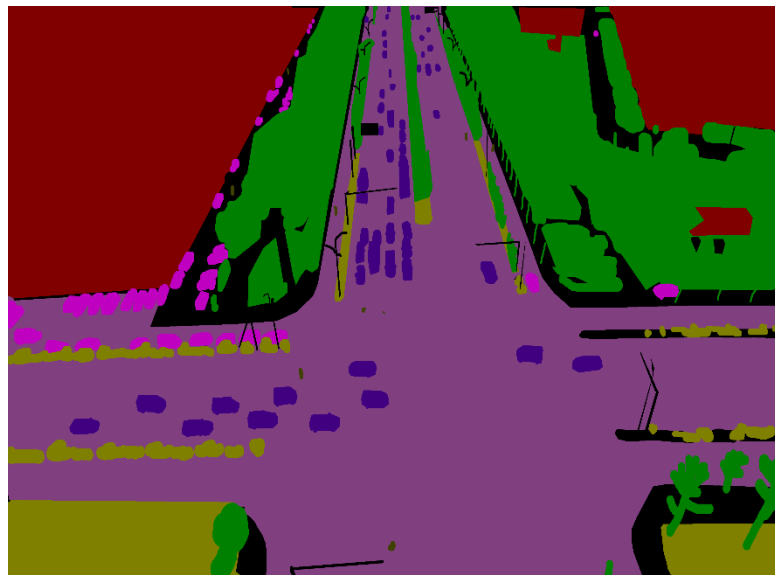
قطعه‌بندی معنایی تصاویر پهلو-مناوای شهری بر مبنای شبکه عصبی رمزگذار-رمزگشا U-Net



کدگذاری تصاویر برچسب

1	name	r	g	b
2	Background	0	0	0
3	Building	128	0	0
4	Road	128	64	128
5	Tree	0	128	0
6	Low_vegeta	128	128	0
7	Moving_car	64	0	128
8	Static_car	192	0	192

Class No.: 1
 Class No.: 2
 Class No.: 3
 Class No.: 4
 Class No.: 5
 Class No.: 6
 Class No.: 7

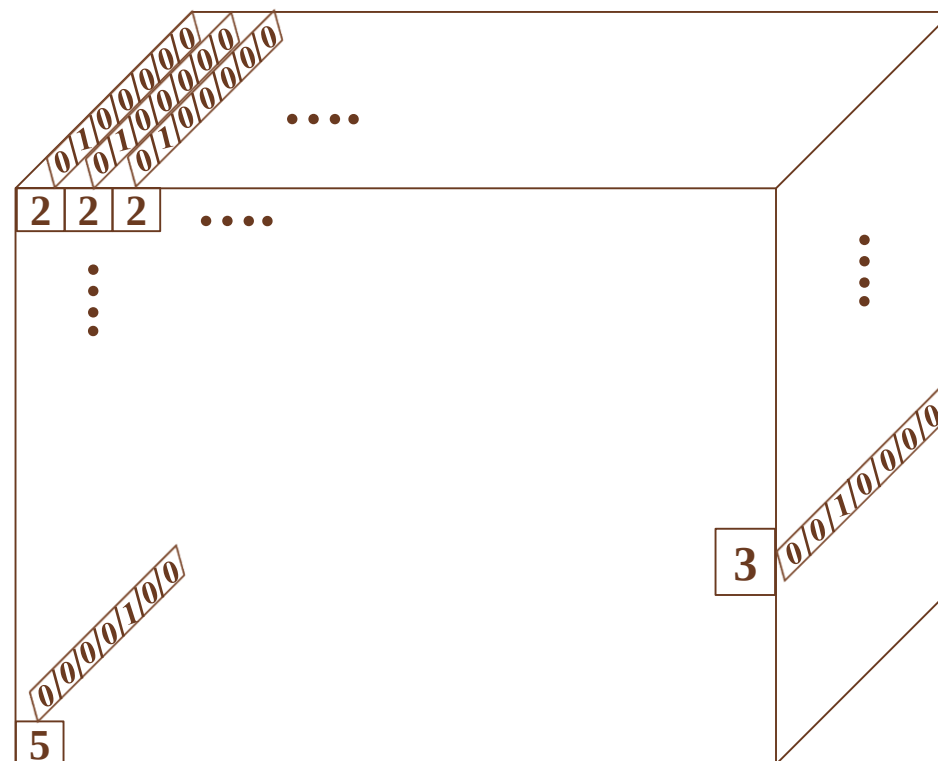




قطعه بندی معنایی تصاویر پهپاد- مبنای شهری بر مبنای شبکه عصبی رمزگذار- رمزگشا U-Net

کدگذاری تصاویر برچسب

1	name	r	g	b
2	Background	0	0	0
3	Building	128	0	0
4	Road	128	64	128
5	Tree	0	128	0
6	Low_vegeta	128	128	0
7	Moving_car	64	0	128
8	Static_car	192	0	192



مقدمه

بیان مسئله و
ضرورت تحقیق

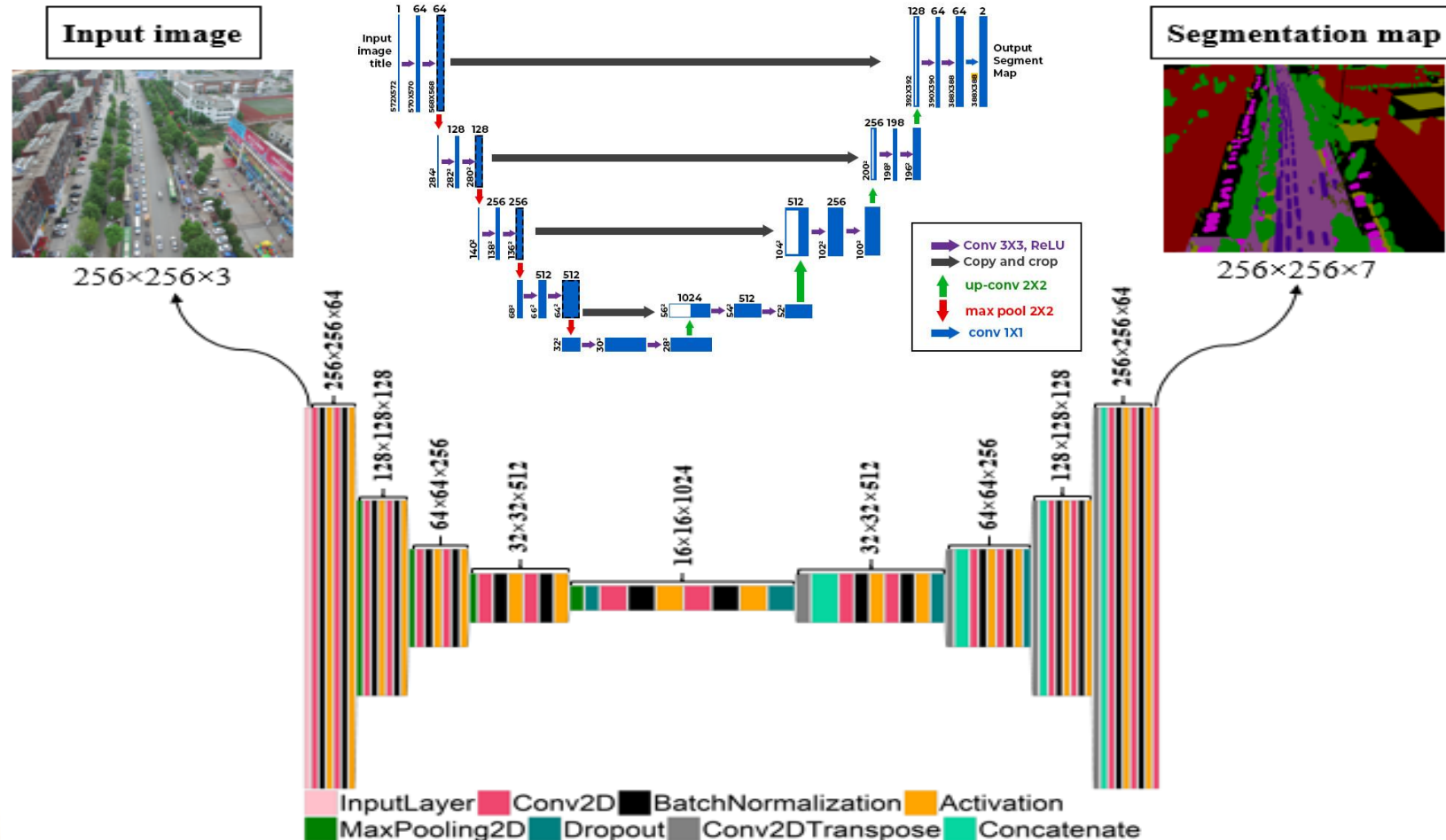
روش شناسی
تحقیق

پیاده سازی ها و
تحلیل نتایج

نتیجه گیری و
پیشنهادها

قطعه بندی معنایی تصاویر پهلو-مناوای شهری بر مبنای شبکه عصبی رمزگذار-رمزگشا U-Net

شبکه عصبی رمزگذار-رمزگشا U-Net



مقدمه

بیان مسئله و
ضرورت تحقیق

روش شناسی
تحقیق

پیاده سازی ها و
تحلیل نتایج

نتیجه گیری و
پیشنهادها

معیارهای ارزیابی

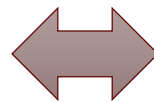
Confusion Matrix		کلاس های واقعی زمینی			
		کلاس ۱	کلاس ۲	...	کلاس ۷
کلاس های پیش بینی شده	کلاس ۱	$C_{1,1}$	$C_{1,2}$	...	$C_{1,7}$
	کلاس ۲	$C_{2,1}$	$C_{2,2}$	...	$C_{2,7}$
	⋮	⋮	⋮	⋮	⋮
	کلاس ۷	$C_{7,1}$	$C_{7,2}$	...	$C_{7,7}$

■ مثبت واقعی و منفی واقعی
■ مثبت کاذب و منفی کاذب

$$F_1 \text{ Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Precision}_i = \frac{C_{ii}}{\sum_{j=1}^{N_{cls}} C_{ij}}$$

$$\text{Recall}_i = \frac{C_{ii}}{\sum_{j=1}^{N_{cls}} C_{ji}}$$



$$\text{Overall Accuracy} = \frac{\sum_{i=1}^{N_{cls}} C_{ii}}{\sum_{i=1}^{N_{cls}} \sum_{j=1}^{N_{cls}} C_{ij}}$$

$$\text{IOU}_i = \frac{\text{target} \cap \text{prediction}}{\text{target} \cup \text{prediction}} = \frac{C_{ii}}{\sum_{j=1}^{N_{cls}} C_{ij} + \sum_{j=1, j \neq i}^{N_{cls}} C_{ji}}$$

$$\text{Mean}_{\text{IOU}} = \frac{1}{N_{cls}} \sum_{i=1}^{N_{cls}} \text{IOU}_i$$

$$\text{MeanBF Score} = \frac{1}{N_{cls}} \sum_{i=1}^{N_{cls}} F_1 \text{ Score}_i$$

مقدمه

بیان مسئله و ضرورت تحقیق

روش شناسی تحقیق

پیاده سازی ها و تحلیل نتایج

نتیجه گیری و پیشنهادها

مجموعه داده مورد مطالعه

UAVid2020



ISPRS
Benchmark

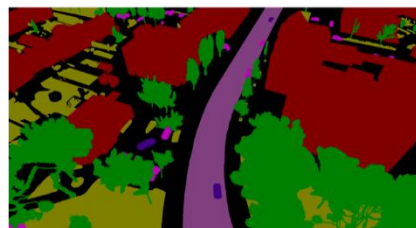
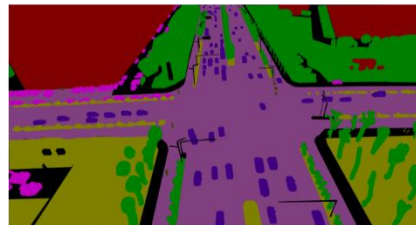
جدول ۱ - جزئیات مجموعه داده اصلی

UAV imaging sensor	DJI Phantom3 Pro	DJI Phantom4
Resolution (width × height) (pixel)	4096 × 2160 / 3840 × 2160 (4K)	
Number of classes	8	

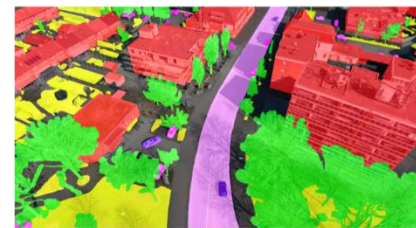
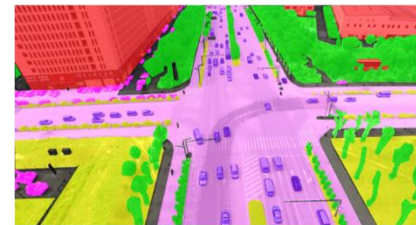
تصویر اصلی



تصویر واقعیت زمینی



تصویر هم‌پوشانی



پس‌زمینه

ساختمان

جاده

درخت

پوشش گیاهی کم

خودرو ثابت

خودرو متحرک

انسان

مقدمه

بیان مسئله و
ضرورت تحقیق

روش‌شناسی
تحقیق

پیاده‌سازی‌ها و
تحلیل نتایج

نتیجه‌گیری و
پیشنهادها

جزئیات پیاده‌سازی



جدول ۲ - تنظیمات مربوط به پارامترهای آموزش

Number of epoch	30
Learning rate	0.001
Optimizer	Adaptive moment estimation (Adam)
Loss Function	Categorical cross entropy

جدول ۳ - مشخصات مجموعه داده مورد آزمایش

Total number of frames	270	The number of training video frames	200
		Number of validation/test video frames	70
Number of classes	7		

مقدمه

بیان مسئله و
ضرورت تحقیق

روش‌شناسی
تحقیق

پیاده‌سازی‌ها و
تحلیل نتایج

نتیجه‌گیری و
پیشنهادها



**دقت قطعه‌بندی معنایی مدل یادگیری عمیق در مقایسه
با روش‌های سنتی یادگیری ماشین**

دقت قطعه‌بندی هر کلاس (%)			
U-Net	MLP	RF	کلاس / روش
۶۳٫۱۶	۲۹٫۶	۳۰٫۴	پس زمینه
۸۲٫۳۲	۷۳٫۱۹	۶۷٫۸۳	ساختمان
۸۱٫۸۱	۴۲٫۵	۴۳٫۰۹	جاده
۵۶٫۱۶	۶۸٫۷۷	۶۲٫۹۳	درخت
۸۱٫۰۶	۳۱٫۰۹	۳۷٫۱۴	پوشش گیاهی کم
۶۶٫۹۱	۰٫۹۷	۱٫۸۴	خودرو متحرک
۲۶٫۸۷	۰	۰٫۷۳	خودرو ثابت

مقدمه

بیان مسئله و
ضرورت تحقیق

روش‌شناسی
تحقیق

پیاده‌سازی‌ها و
تحلیل نتایج

نتیجه‌گیری و
پیشنهادها



مقادیر معیار IOU برای هر کلاس

مقادیر IOU (%)			
U-Net	MLP	RF	کلاس / روش
۴۰٫۵۴	۲۰٫۹۹	۲۰٫۲۶	پس زمینه
۷۶٫۲۷	۴۷٫۵۸	۴۵٫۳۰	ساختمان
۶۰٫۴۰	۲۵٫۷۷	۲۵٫۷۸	جاده
۵۰٫۳۵	۴۵٫۸۴	۴۲٫۷۸	درخت
۴۹٫۳۴	۲۷٫۳۸	۲۸٫۶۲	پوشش گیاهی کم
۴۴٫۴۳	۰٫۹۵	۱٫۶۲	خودرو متحرک
۲۱٫۲۳	۰	۰٫۶۳	خودرو ثابت

مقدمه

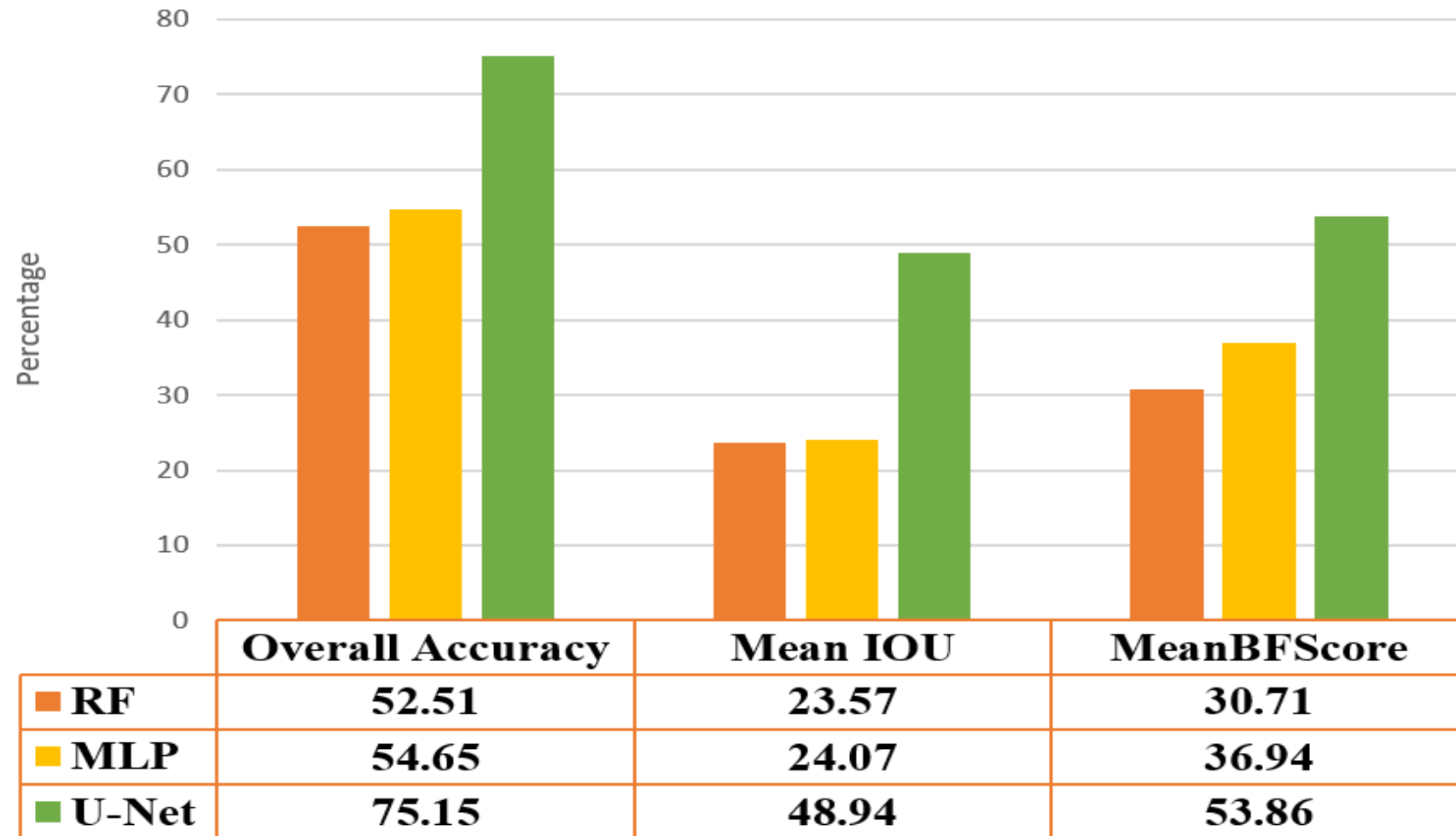
بیان مسئله و
ضرورت تحقیق

روش شناسی
تحقیق

پیاده سازی ها و
تحلیل نتایج

نتیجه گیری و
پیشنهادها

قطعه‌بندی معنایی تصاویر پهلو-بنا نواحی شهری بر مبنای شبکه عصبی رمزگذار-رمزگشا U-Net



مقدمه

بیان مسئله و
ضرورت تحقیق

روش‌شناسی
تحقیق

پیاده‌سازی‌ها و
تحلیل نتایج

نتیجه‌گیری و
پیشنهادها

قطعه‌بندی معنایی تصاویر پهلو-بنا نواحی شهری بر مبنای شبکه عصبی رمزگذار-رمزگشا U-Net



نتایج بصری

مقدمه

بیان مسئله و
ضرورت تحقیق

روش‌شناسی
تحقیق

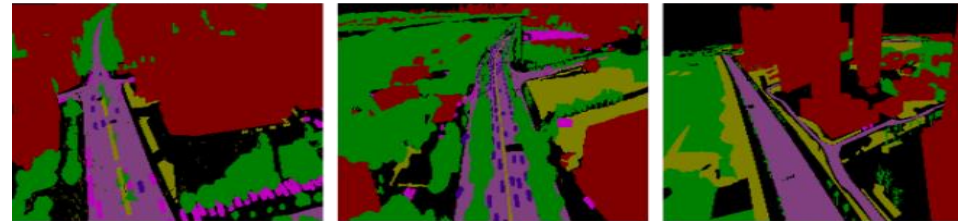
پیاده‌سازی‌ها و
تحلیل نتایج

نتیجه‌گیری و
پیشنهادها

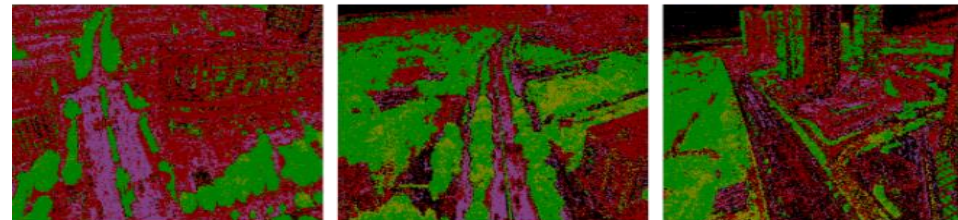
Original Image



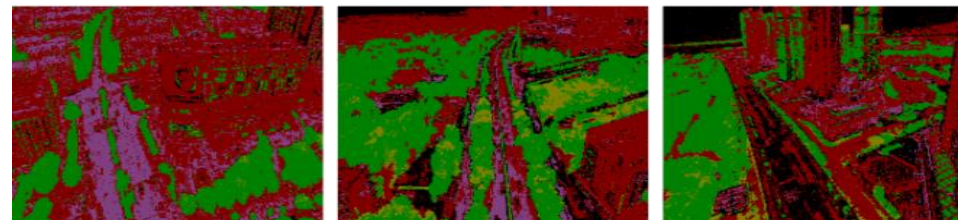
Ground Truth Image



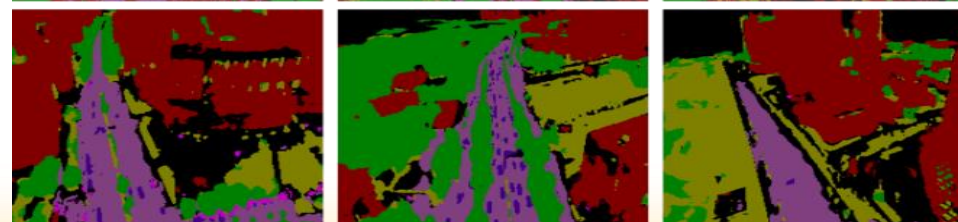
Random Forest



Multi-Layer Perceptron



U-Net



پس زمینه	ساختمان
جاده	درخت
پوشش گیاهی کم	
ماشین متحرک	ماشین ثابت



نتیجه‌گیری

مقدمه

بیان مسئله و
ضرورت تحقیق

روش‌شناسی
تحقیق

پیاده‌سازی‌ها و
تحلیل نتایج

نتیجه‌گیری و
پیشنهادها

- ✓ نتایج آماری و بصری به دست آمده، عملکرد بالای روش یادگیری عمیق را در قطعه‌بندی معنایی نواحی شهری مبتنی بر پهپاد نشان می‌دهد.
- ✓ برتری مدل یادگیری عمیق U-Net در مقایسه با الگوریتم‌های یادگیری ماشین سنتی جنگل تصادفی (RF) و پرسپترون چند لایه (MLP).
- ✓ مدل قطعه‌بندی معنایی (شبکه رمزگذار-رمزگشا کانولوشنی U-Net) در برخورد با چالش‌هایی نظیر وجود اشیا یکسان با مقیاس‌های متفاوت در فریم‌های ویدئویی و تشخیص اشیاء متحرک از ساکن، عملکرد مطلوبی ارائه داد.



پیشنهادهای

1. قطعه‌بندی معنایی مجموعه داده پهپاد-مبنای مورد مطالعه، در ابعاد و وضوح اصلی انجام شود.
2. در این مطالعه برای قطعه‌بندی معنایی از شبکه یادگیری عمیق رمزگذار-رمزگشا کانولوشنی استفاده شد، که در مطالعات بعدی می‌توان توانایی شبکه‌های عصبی دیگر نظیر شبکه‌های عصبی باقی‌مانده، شبکه‌های مولد تخصصی، مدل‌های ترنسفورمر و غیره را نیز بررسی کرد.
3. استفاده از رویکرد یادگیری انتقالی برای تقسیم‌بندی معنایی تصاویر هوایی شهری.

مقدمه

بیان مسئله و
ضرورت تحقیق

روش‌شناسی
تحقیق

پیاده‌سازی‌ها و
تحلیل نتایج

نتیجه‌گیری و
پیشنهادهای

PHASEONE



استخراج اتوماتیک ساختمان از
تصاویر هوایی اورتو شهر اهواز



- [1] P. Sturgess, K. Alahari, L. Ladicky, and P. H. Torr, "Combining appearance and structure from motion features for road scene understanding," in *BMVC-British Machine Vision Conference*, 2009: BMVA.
- [2] P. Kotschieder, S. R. Buló, H. Bischof, and M. Pelillo, "Structured class-labels in random forests for semantic image labelling," in *2011 international conference on computer vision*, 2011: IEEE, pp. 2190-2197.
- [3] T. Moranduzzo and F. Melgani, "A SIFT-SVM method for detecting cars in UAV images," in *2012 IEEE International Geoscience and Remote Sensing Symposium*, 2013: IEEE, pp. 6868-6871.
- [4] Y. Lyu, G. Vosselman, G.-S. Xia, A. Yilmaz, and M. Y. Yang, "UAVid: A semantic segmentation dataset for UAV imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 165, pp. 108-119, 2020.
- [5] Y. Pi, N. D. Nath, and A. H. Behzadan, "Detection and semantic segmentation of disaster damage in UAV footage," *Journal of Computing in Civil Engineering*, vol. 35, no. 2, p. 04020063, 2021.

با سپاس فراوان از توجه شما

ودمیان سلام و نور مهربانی خداوند بر شما تابنده تر از همیشه....