ScienceDirect

Cities
Volume 96, January 2020, 102481

Understanding cities with machine eyes: A review of deep computer vision in urban analytics

Mohamed R. Ibrahim, James Haworth, Tao Cheng

Show more

Outline | Share Cite

https://doi.org/10.1016/j.cities.2019.102481
Get rights and content

Highlights

- The basics and algorithms of computer vision.
- The role of deep learning and computer vision in tackling complexity in cities.
- The application of computer vision in understanding cities to-date.
- Challenges and opportunities for relying on computer vison in tackling urban issues.
- Recommendations towards AI-generated urban policies.

Abstract

Modelling urban systems has interested planners and modellers for decades. Different models have been achieved relying on mathematics, cellular automation, complexity, and scaling. While most of these models tend to be a simplification of reality, today within the paradigm shifts of artificial intelligence across the different fields of science, the applications of computer vision show promising potential in understanding the realistic dynamics of cities. While cities are complex by nature, computer vision shows progress in tackling a variety of complex physical and non-physical visual tasks. In this article, we review the tasks and algorithms of computer vision and their applications in understanding cities. We attempt to subdivide computer vision algorithms into tasks, and cities into layers to show evidence of where computer vision is intensively applied and where further research is needed. We focus on highlighting the potential role of computer vision in understanding urban systems related to the built environment, natural environment, human interaction, transportation, and infrastructure. After showing the diversity of computer vision algorithms and applications, the challenges that remain in understanding the integration between these different layers of cities and their interactions with one another relying on deep learning and computer vision. We also show recommendations for practice and policy-making towards reaching AI-generated urban policies.

Previous

Next

Keywords

Cities; Computer vision; Deep learning; Convolutional neural networks (CNN); Urban studies

1. Introduction

Cities are complex entities by nature and modelling urban systems has interested planners for decades (Batty, 2008; Bettencourt, 2013; Isalgue et al., 2007). A range of approaches have been used to model urban processes, examples of which include cellular automata (Batty, 1997; Batty, Couclelis, & Eichen, 1997; de Almeida et al., 2003), fractals (Batty and Longley, 1994; Batty and Xie, 1996; Frankhauser, 1998; Murcio et al., 2015) and multi-agent models (Batty, 2005; Heppenstall, Crooks, See, & Batty, 2012). These models aim to understand cities by modelling their underlying components and exploring their systems, ultimately intending to inform decision making and policy (Batty, 2009; Calder et al., 2018). Due to the complexity and nonlinearity of cities, these models tend to explore or predict urban systems in a sectoral fashion. For example, transport models are used to simulate the potential impact of policy and infrastructure investment. Such models may fail to represent complex events in cities, in which multiple systems interact.

The success of deep learning and computer vision in pattern recognition over the past decade (LeCun et al., 2015) has created opportunities to understand cities through images (Reichstein et al., 2019). So far, the diversity of the algorithms of computer vision has enabled researchers to tackle and predict a wide spectrum of issues in more accurate and precise fashion (Goodfellow et al., 2017; LeCun et al., 2015; Reichstein et al., 2019).

In this paper, we review the algorithms and applications of computer vision related to urban analytics. Urban analytics can be defined as urban research that exploits new data resources that are captured, for example, from sensors (e.g. imagery, the internet of things), crowdsources and social media (Batty, 2019). Deep learning and computer vision technologies have tremendous potential in this area for dealing with heterogeneous data types, many of which are image-based. In the review, we identify the areas that have been intensively modelled using computer vision while also revealing the areas in which further research is needed. This is achieved by categorising the application areas of urban analytics into five layers of the city (the built environment, human interaction, transportation and traffic, the natural environment, and infrastructure). In doing so, we demonstrate that, while many urban processes are a result of interactions across these layers, the current approach is to tackle these layers differently and separately. Here, we note the potential of extracting data of different disciplines using a unified input (images/videos) that relies on computer vision methods to cover a wide spectrum of urban and transport research.

This review aims to provide a resource for urban planners and practitioners by: 1) reviewing the main methodologies of computer vision, and their applicability to various tasks of urban analytics, 2) illustrating the variation and nuances of deep learning and computer vision algorithms and their limitations in understanding cities, 3) giving a descriptive understanding of the algorithms of computer vision for policy-makers and planners, and how they are used in cities, 4) paving the way for developing AI-generated urban policies by highlighting the key enabling technologies and research directions. The remainder of this review is structured as follows: In Section 2, the methodology of the review is described. In Section 3, the key tasks of computer vision are described, along with the main algorithms. The applications of computer vision in urban analytics are reviewed in Section 4. Section 5 summarises what remains missing in current research, before Section 6 shows how we can move from prediction to decision making and policy recommendation. Finally, some conclusions are given in Section 7.

2. Review methodology

The methodology of this review is divided into two parts: 1) manuscripts are collected that summarise the progress in deep learning methods and algorithms that are applicable to computer vision tasks, 2) manuscripts are collected that reflect the application of deep learning and computer vision in understanding cities in the last decade (since 2010). For the first part, we present only the major methodological approaches. Papers that vary or improve on these main approaches are excluded. Most of these studies are presented in premier computer science conferences, including, but not limited to CVPR, ICCV, ECCV and NeurIPS, or in ArXiv. For the second part, we extend the search to peer-reviewed journals and conference proceedings listed in Scopus, Web of Science, Google Scholar and Science Direct, that can be accessed via a combination of keywords such as: deep learning, cities, computer vision, land-use modelling, urban perception, prediction, detection, street-level images, aerial or satellite images. This is because the applied computer vision literature is often found in domain specific journals, rather than computer science conferences.

In total, 641 manuscripts were collected to cover the two parts of the methodology. For the second part, the collected manuscripts were filtered to include only those related to computer vision of street-level or aerial images, which use deep learning or hybrid models that include a convolutional structure. Studies that involve deep learning of other data types such as 2D/3D LIDAR data are excluded. Studies that use classical machine learning or computer vision algorithms without involving deep learning are also excluded, except where they are required to draw a baseline to emphasise advancement or contrast. The algorithms are presented at a descriptive level and readers are referred to the relevant literature for further details.

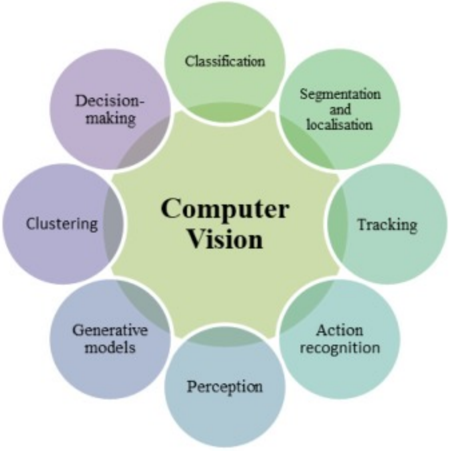
3. The Basics and tasks of computer vision

Before exploring the domains where computer vision is applied in cities, it is worth identifying first what computer vision is and what its algorithms are capable of achieving from a generic perspective. Computer vision can be narrowed to the task of learning the qualitative representation of visual elements in their raw form in order to quantify them (LeCun et al., 2015). Similar to human eyes, the computer sees visual objects and creates a cognitive understanding of a scene based on a sequential sample of the presented images or frames of images in a task-specific manner. While computer vision is not new (i.e. Viola & Jones, 2001), deep learning, most specifically Convolutional Neural Networks (CNN), has made it possible for computer vision to tackle various issues and process images more precisely and efficiently (He et al., 2015; LeCun et al., 2015). These deep models, computation capabilities, and the availability of large datasets have made it possible for computer vision to permeate a wide range of applications in realistic settings (Cordts et al., 2016; Lin et al., 2014; Russakovsky et al., 2015). Generally, the logic of computer vision, relying on these deep models, can be summarized as the construction of multiple hidden layers that are capable of accomplishing a range of vision tasks by extracting digital features that may or may not be recognisable to human eyes (Guo et al., 2016; Kuo, 2016; LeCun et al., 2015). The most commonly used are convolutional, pooling, flatten, and fully-connected layers. The general functions of these layers can be summarised as follows:

- Convolutional layers are responsible for extracting features coupled with activation functions, such as Rectified linear units (ReLU), to add nonlinearity to the model,
- Pooling layers are responsible for reducing the dimensionality of the data,
- Flatten layers are responsible for converting the features of the model into neurons to be fed forward to the fully-connected layers
- Fully-connected layers aim to adjust the weights and predict the output for a given task.

The types, numbers, and orders of these layers are responsible for determining functionality and the optimisation of both accuracy and time needed for the training and the inference of the model. The structure of the model and the fine-tuning of the various hyperparameters represents the innovation and the advancements of the state-of-the-art for pattern recognition for a given task (LeCun et al., 2015).

Depending on the type of visual task, deep models can be trained differently with different layers and different sets of algorithms (Guo et al., 2016). As shown in Fig. 1, these algorithms of computer vision can be subdivided based on eight fundamental tasks, upon which other tasks can be framed and built. These are; image classification, segmentation and localisation, tracking, action-recognition, perception, generative models, clustering, and decision-making. Table 1 shows the literature related to different computer vision tasks. It expands on the methods related to each task and their subcategories.



Download : Download high-res image (151KB)

Download : Download full-size image

Fig. 1. Computer vision tasks. The figure is created by the authors.

Table 1. Methods related to the tasks of computer vision.

Vision task	Sub-category	method	
Classification		AlexNet	(**Krizhevsky et al., 2012**)
		VGGNet	(simonyan & zisserman, 2014)
		GoogLeNet	(Szegedy et al., 2015)
		ResNet	(he et al., 2015)
		DensNet	(huang, liu, weinberger, & van der maaten, 2017)
Segmentation and localisation	**Object-based detection**	R-CNN	(Girshick et al., 2014)
		Fast R-CNN	(Ren et al., 2016)
		YOLO	(Redmon et al., 2016)
		SSD	(Liu et al., 2016)
		YOLOv2	(Redmon & Farhadi, 2017)
		YOLOv3	(Redmon & Farhadi, 2018)

Vision task	**Sub-category**	**method**	
		RetinaNet	(Lin et al., 2018)
	Semantic segmentation	**DeepLab**	(Chen, Papandreou, Kokkinos, Murphy, & Yuille, 2016)
		U-Net	(Ronneberger et al., 2015)
		SegNet	(Badrinarayanan et al., 2016)
		–	(Long et al., 2015)
		–	(Peng et al., 2017)
		–	(Chen, Papandreou, Kokkinos, Murphy, & Yuille, 2016)
		–	(Zhao et al., 2017)
		–	(Yu & Koltun, 2015)
		RefineNet	(Lin et al., 2017)
		–	(Chen et al., 2017)
		–	(Jégou et al., 2016)
		FoveaNet	(Li, Wang, et al., 2017)
		LinkNet	(Chaurasia & Culurciello, 2017)
		–	(Yang et al., 2018)
Tracking objects		MOTS	**(Voigtlaender et al., 2019)**
		–	(Jiang et al., 2018)
		–	(Kang et al., 2016)
		–	(Girdhar et al., 2017)
		–	(Danelljan et al., 2015)
		–	(Held et al., 2016)
		ECO	(Danelljan et al., 2016)
		CNNTracker	(Chen et al., 2016)
		ArtTrack	(Insafutdinov, Andriluka, et al., 2016)
		–	(Wu et al., 2016)
		–	(Chu et al., 2017)
		PathTrack	(Manen et al., 2017)
Action recognition	**Human pose estimation**	DensePose	(Guler et al., 2018)
		MultiPoseNet	(Kocabas et al., 2018)
		–	(Papandreou et al., 2017)
		RMPE	(Fang et al., 2016)
		DeeperCut	(Insafutdinov, Pishchulin, et al., 2016)
		–	(Cao et al., 2016)
		–	(Pfister et al., 2015)
	Action classification	–	(Girdhar & Ramanan, 2017)
		–	(Bilen et al., 2016)
		–	(Zhu, Vial, et al., 2017; Zhu, Lan, et al., 2017)
		–	(Guo, Chou, et al., 2018; Guo, Huang, et al., 2018)
		–	(Zhang, Wang, et al., 2016; Zhang, Xu, et al., 2016)
	Temporal action detection	**Daps**	(Escorcia et al., 2016)
		–	(Diba et al., 2017)
		–	(Gemert, Jain, Gati, & Snoek, 2015)

Vision task	Sub-category	method	
Perception	Spatio-temporal action detection	–	(Shou et al., 2017)
		–	(Escorcia et al., 2016)
		–	(Li et al., 2016)
		–	(Xu et al., 2017)
		–	(Chao et al., 2018)
		–	(Buch et al., 2017)
		–	(Zhao, Shi, et al., 2017; Zhao, Xiong, et al., 2017; Zhou, Zhao, et al., 2017)
		–	(Chen & Corso, 2015)
		–	(Becattini et al., 2017)
		–	(Saha et al., 2017)
		–	(Gemert et al., 2015)
		–	(Zhu, Vial, et al., 2017; Zhu, Lan, et al., 2017)
		–	(El-Nouby & Taylor, 2018)
		–	(Saha et al., 2016)
		–	(Singh et al., 2016)
		–	(Mettes et al., 2016)
		–	(Weinzaepfel et al., 2015)
	Understanding scenes		(Eslami et al., 2018)
	Estimating depth	–	(Cao et al., 2017)
		–	(He et al., 2018)
Generative models	**GANS**	–	(Goodfellow et al., 2014)
		–	(Radford et al., 2015)
		–	(Reed, Akata, Yan, et al., 2016; Reed, Akata, Mohan, et al., 2016)
		StackGAN	(Zhang, Wang, et al., 2016; Zhang, Xu, et al., 2016)
		–	(Isola et al., 2016)
		BigGAN	(Brock et al., 2018)
Clustering		–	**(Caron et al., 2018)**
		–	(Xie et al., 2016)
		DeepCluster	(Tian et al., 2017)
Making decisions	**Deep Q-learning**	–	(Mnih et al., 2013)
		–	(Hester et al., 2017)
	Double Deep Q-learning	–	(van Hasselt et al., 2015v)
	Duel Deep Q-learning	–	(Wang, Qiao et al., 2015; Wang, Schaul, et al., 2015)
	A3C	–	(Mnih et al., 2016)

3.1. Classification

Deep learning models, most specifically Convolutional Neural Networks (CNN), have shown substantial progress in classifying images of a wide spectrum of classes (LeCun et al., 2015). Various deep CNN models with different architectures and hyper-parameters have been computed to recognize visual objects in large repositories of images, such as the ImageNET dataset that contains 15 million images that belong to 22,000 different classes (Russakovsky et al., 2015, 2015). Starting with AlexNet (Krizhevsky et al., 2012), VGGNet (Simonyan & Zisserman, 2014), GoogLeNet (Szegedy et al., 2015), ResNet (K. He et al., 2015) and most recently, DenseNet (Huang et al., 2017), these CNN models are able to accurately recognize and classify a wide range of images. For instance, ResNet-152 achieved 4.49% top-5 error score on the validation set of ImageNET (K. He et al., 2015).

3.2. Segmentation and localisation

Segmentation and localisation are the processes of identifying multiple objects in a single image. These models use a single deep model in an end-to-end fashion, in which the first part of the model is an image classifier followed by different types of layers to localise different objects with a given confidence. Notable examples include the Region-based CNN model (R-CNN) (Girshick et al., 2014), Fast R-CNN (Ren et al., 2016), You Only Look Once (YOLO) (Redmon & Farhadi, 2017, 2018) and the MultiBox Detectors for fast image segmentation, or so-called; Single Shot Multi-Box Detector (SSD) technique (Liu, Tsow, et al., 2016; Liu, Anguelov, et al., 2016). CNN models have shown significant progress in recognising and detecting objects in images with a minimal inference time and high overall validation accuracy. YOLOv3 achieves 93.8% top-5 score on the COCO dataset (Redmon & Farhadi, 2018).

For further explanation related to localisation and object detection, see (Zou et al., 2019).

3.3. Tracking objects

After building a system of object detection, computer vision can be used for tracking multiple objects in a complex scene by adding features that correlate a pair of consecutive frames. This tracker system is capable of identifying a candidate box at each frame-level jointly with their time deformations (Girdhar et al., 2017). While different tracker systems can be built based on correlation filtering and online learning techniques between consecutive frames (Zhang, Xia, et al. (2018); Zhang, Zhou, et al. (2018)), the state-of-the-art research in object tracking uses an end-to-end CNN model to tackle both detection and tracking, which can add more advanced features (i.e. dealing with occlusion issues) for tracking various elements (Girdhar et al., 2017; Hou et al., 2017; Kang et al., 2016). For further explanation related to deep visual tracking, see P. Li, Wang, Wang, & Lu (2018).

3.4. Action recognition

Computer vision coupled with deep CNN models is not only capable of tracking the motion of an object in a complex scene, but also classifying its multiple actions while tracking (Bilen et al., 2016; Limin Wang, Qiao et al., 2015; Wang, Schaul, et al., 2015; Zhang, Wang, et al., 2016; Zhang, Xu, et al., 2016). Various computer vision algorithms have been developed to tackle humans poses and their interaction with an external object in a complex scene (El-Nouby & Taylor, 2018; Saha et al., 2016; Soomro & Shah, 2017; Weinzaepfel et al., 2016). 2D or 3D convolution layers (with or without the spatiotemporal dimensions) can identify the action of the object from its pose in relation to another target object. For instance, from the pose of a person sitting on a bike, the algorithms of computer vision can identify cycling as an action. This concept of the triplet inputs (object, verb, target) has been seminal for tackling real-world events and behaviours, from a simple still image to multi-frame images (Girdhar et al., 2017).

3.5. Perception

Perception tasks can be seen as classification or regression tasks that predict information that is not necessarily embedded directly in the image but can be inferred from the overall structure of the image. Perceiving a neighbourhood as safe or unsafe for example can be seen as a perception task, in which the machine extracts features from the structure of an image to classify the safety of the image. Even though understanding the overall gist of a scene is seminal for understanding more than an object in an image (Oliva & Torralba, 2006), few works have been done in this domain. The complexity of tackling this subject lies in sensing the class of an image by sensing the overall profound features of the image, rather than identifying an object in the image. For instance, identifying and sensing the planning status of a region from the image (Ibrahim et al., 2019).

Moreover, seeing what is far and what is close just by looking at a still image is another advantage of computer vision relying on deep CNN models. Cao, Wu, & Shen (2017) trained deep CNN models to estimate the depth in a single image by labelling the different depths on the image and dealing with training the model as a classification task. In contrast, He, Wang, & Hu (2018) trained a deep CNN model to estimate the depth of a monocular image relying on the information of focal length that has proven to outperform the other state-of-the-art depth estimation algorithms based on deep learning models.

3.6. Generative models

Generative models refer to the ones that tend to output synthesized data by learning the representation of their input data in an unsupervised fashion, conditionally or unconditionally.

There is a range of algorithms that are classified as generative models, such as Restricted Boltzmann Machine (RBM), deep belief networks, Autoencoders, and Generative adversarial Networks (GANs) (Goodfellow et al., 2017). This section refers only to GANs, which generate synthetic graphical data in an unsupervised training fashion relying on images as input. Unlike other tasks related to computer vision, the deep models of GANs, introduced in 2014, enable machines to generate new information that is similar to what the model has been trained to identify (Goodfellow et al., 2014). In other words, if the model is trained on images of trees, by using GANs the model can generate a new image of a tree that preserves the fundamental features of a tree, but with a new visual identity. This progress of deep learning enables the creation of unique objects or scenes by understanding the underlying features of the trained images or videos.

GANs are trained differently from the abovementioned deep models, not only in term of layers but rather, instead of the single end-to-end model, two parallel deep models are trained that compete with one another (Goodfellow, 2016; Goodfellow et al., 2014; Radford et al., 2015). The first one, the Generator model, generates new images to deceive the second model that holds the ground truth data, while the second model, the Discriminator model, blocks this new image until the generator model becomes advanced enough to generate new images that are similar enough to the ground truth that the discriminator model can no longer refuse them. This computationally intensive training, in an unsupervised manner, opens the door for computer-based creativity without the prior supervision of humans.

GANs have been utilised in various applications. Isola, Zhu, Zhou, & Efros (2016) used conditional GANs to translate from one form of an image to another. For instance, by giving the model a satellite image of a location, the model can give the semantic segmentation of the location or vice versa. Zhang, Wang, et al. (2016); Zhang, Xu, et al. (2016) created stackGAN model to transform a text description of an image into a photo-realistic synthesis. Moreover, Reed, Akata, Yan, et al. (2016); Reed, Akata, Mohan, et al. (2016) have pushed the algorithms of GANs further. The machines can learn to draw not only from text distributions but also by telling the machine what and where to draw on the canvas. Apart from the daily-life applications, GANs have been used in the simulation of 3D energy particle showers and physics-related applications (Paganini et al., 2018).

3.7. Clustering

Clustering is a form of unsupervised learning, in which the machines are able to cluster different still images or multi-frame images based on their content or embedded objects without prior human supervision (Caron et al., 2018; Tian et al., 2017; Xie et al., 2016). So far, different computer vision algorithms have been developed to tackle this task and eliminate the need for a long process of manual labelling from still images. Recently, Eslami et al. (2018) introduced the Generative Query Network (GQN) for scene representation without human supervision. The GQN takes images from a different perspective as an input and generates a visual representation of the scene from an unobserved perspective. This process of coupling generative models with clustering introduces a new form of machine intelligence to understand scene representation without human supervision.

3.8. Decision-making

By looking at the edge of computer vision and coupling its deep models with reinforcement learning, or so-called Deep Reinforcement Learning (DRL), machines can be trained to explore and compute the outcomes of different scenarios in order to make real-time decisions based on visual aspects of the environment (Hester et al., 2017; Mnih et al., 2016). This level of cognitive ability of machines by applying one or more of the abovementioned tasks can enable an agent to grasp information and interact with an environment to optimize target resources without human supervision.

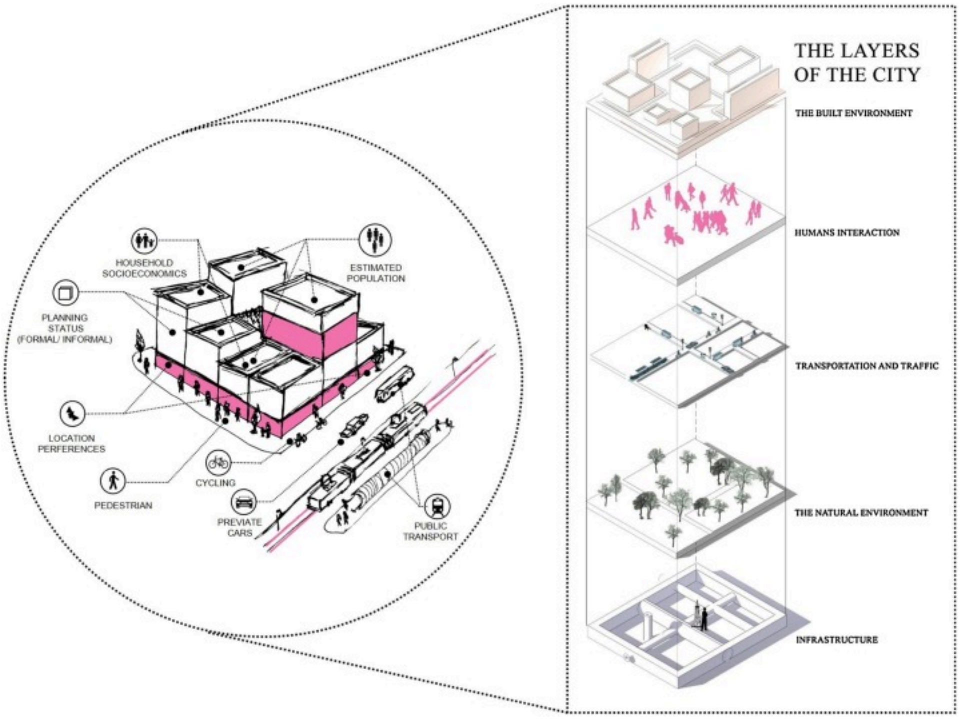
Due to the complexity of the algorithms related to this subject, most examples are in virtual or gaming environments (Mnih et al., 2013). However, most significantly, Mirowski et al. (2018) utilised DRL to enable a machine to navigate through the unstructured environment of the street network relying on street-level images. In this work, the machine learns to navigate by understanding landmarks from images and to determine its location and its target destination.

4. Recognising the urban world

Understanding the dynamics of cities remains a complex issue. Data collection, for instance, is one of the crucial domains where automation is highly desirable, in which computer vision has been successfully applied in capturing and analysing various objects depicted in urban scenes. Specifically, scene parsing and semantic segmentation represent crucial tasks of computer vision for a better understanding of the elements of an urban scene. From images, computer vision can localize multiple objects in cities, or simply segment the entire scene based on a group of themes, such as sky, ground, road, building, vegetation, etc (Chaurasia & Culurciello, 2017; Zhou et al., 2017).

Putting all the above-mentioned tasks together, computer vision shows good potential in urban analytics for analysing the multi-layers of cities. For the purposes of this review, we define these layers as; the built environment, the natural environment, humans and their physical interactions, transport modes and traffic-related issues, and infrastructure. The main reason for breaking-down cities in these layers is to be able to tackle the applications of computer vision in each individual field of science related to urban analytics, in which the methods, scope, language used, and the nature of work may vary depending on the discipline. For instance, research that has been done in understanding the built environment may vary in nature from that done to understand transportation, even though the methods of deep learning and computer vision may be similar.

Fig. 2 shows examples of computer vision applications in cities to detect multidisciplinary tasks that belong to the five layers of cities, whereas Table 2 shows the applications of computer vision to these layers. Each layer is broken down into further subcategories as appropriate.



Download : Download high-res image (664KB)

Download : Download full-size image

Fig. 2. The layers of the cities where computer vision has been applied. The figure is created by the authors.

Table 2. Computer vision algorithms that tackle urban-related issues.

City layer	Category	Method	
The built environment	**Urban components**	Semantic segmentation	(Zhou et al., 2017)
			(**Chaurasia & Culurciello, 2017**)
			(**Chen, Dowman, et al., 2016**; **Chen, Yang, et al., 2016**)
			(**He et al., 2019**)
			(**Helbich et al., 2019**)
			(**Amirkolaee & Arefi, 2019**)
			(**Wurm et al., 2019**)
			(**Cordts et al., 2016**)
		Object-based detection	(**Yang et al., 2019**)
			(**Chew et al., 2018**)
	Land Use classification	**Classification and semantic segmentation**	(Demir et al., 2018)
			(**Sharma et al., 2017**)
			(**Audebert et al., 2018**)
		classification	(**Wang, Zhou, et al., 2018**; **Wang, Xu, et al., 2018**; **Wang, Quan, et al., 2018**; **Wang, Yang, et al., 2018**)
			(**Srivastava et al., 2019**)
			(**Chew, Amer, et al., 2018**; **Chew, Jones, et al., 2018**)
	Urban perception	**Classification and perception**	(Ibrahim et al., 2019)
			(**Zhao et al., 2018**)
			(**Law, Seresinhe, et al., 2018**)
			(**Zhang et al., 2019**)
			(**Seresinhe et al., 2017**)

City layer	Category	Method	
			(**Oliva & Torralba, 2006**)
			(**Wang, Zhou, et al., 2018**; **Wang, Xu, et al., 2018**; **Wang, Quan, et al., 2018**; **Wang, Yang, et al., 2018**)
			(**Salesses et al., 2013**)
			(**Dubey et al., 2016**)
			(**Naik et al., 2016**)
			(**Quercia et al., 2014**)
			(**De Nadai et al., 2016**)
Human interaction	Urban safety		(**Naik et al., 2014**)
		Object-based detection	(Priya et al., 2015)
Transportation and traffic	**Traffic surveillance**	Classification and Object-based detection	(Bottino et al., 2016)
		Action recognition	(Yu et al., 2017)
		Object-based detection	(Yang & Pun-Cheng, 2018)
	Safety/ accidents	**Classification and object-based detection**	(Sayed et al., 2013)
			(**Zaki et al., 2013**)
The natural environment	**Flora and fauna**	Object-based detection	(Cai et al., 2018)
			(**Hong et al., 2019**)
		Semantic segmentation	(**Krause et al., 2018**)
			(**Williams et al., 2017**)
		Classification	(**Mohanty et al., 2016**)
			(**Sun et al., 2017**)
	Environmental and weatherconditions	**Classification andperception**	(Liu, Racah, et al., 2016)
			(**Liu, Yang, et al., 2017**)
			(**Guerra et al., 2018**)
			(**Elhoseiny et al., 2015**)
			(**Sirirattanapol et al., 2019**)
Infrastructure	**Concrete condition**	Object-based detection	(Cha et al., 2017)
			(**Wang, Zhou, et al., 2018**; **Wang, Xu, et al., 2018**; **Wang, Quan, et al., 2018**; **Wang, Yang, et al., 2018**)
	Pavement/ road condition	Object-based detection	(Maeda et al., 2018)
	Bridge component recognition	Semantic Segmentation	(Narazaki et al., 2017)

4.1. The built environment

This section addresses cities from an architectural and urban design perspective, for example, understanding cities from a land-use perspective, the level of the physical appearance of the street-level that may indicate or measure housing prices, or even the level of safety with a certain neighbourhood.

When it comes to understanding the built environment, there are different challenges that face urban planners and policy-makers. For example, modelling the physical appearance of complex urban areas is a multi-faceted issue that is vital for planners and policy-makers for making decisions for improving living conditions in cities. The collection of data that reflects the current status of the built environment is a critical issue for urban analytics. So far, the applications of computer vision have merged not only to detect various urban components but also to understand the appearance and the safety factors of an urban scene. While there is a wide range of applications of computer vision in cities, these applications can be divided into two approaches that either analyse cities from street-level images or remote sensing data such as satellite images.

4.1.1. Seeing cities from above

Analysing cities from above relying on remote sensing and geographical information systems (GIS), perhaps, is the most common approach for planners (Chen, Dowman, et al., 2016; Chen, Yang, et al., 2016). Applications of computer vision jointly with these systems are capable of automating urban tasks such as mapping and zoning. Most recently, the notion of DeepGlobe (Demir et al., 2018) aimed to describe the earth from satellite images. DeepGlobe can extract streets, buildings and the different types of land-cover. Similarly, Wang, Zhou, et al. (2018); Wang, Xu, et al. (2018); Wang, Quan, et al. (2018); Wang, Yang, et al. (2018) used a CNN model to segment satellite images into multi-classes at the pixel level. Marcos, Volpi, Kellenberger, & Tuia (2018) used the CNN model for land cover mapping, solving the issue of rotation of objects. Vanhoey et al. (2017) introduced VarCity as an approach of automating the construction of a city-scale 3D model based on semantic segmentation and machine processing of urban components (buildings, built environment, vegetation, roads, etc).

Furthermore, relying on deep learning, Amirkolaee & Arefi (2019) estimated heights from single aerial images, Wang, Zhou, et al. (2018); Wang, Xu, et al. (2018); Wang, Quan, et al. (2018); Wang, Yang, et al. (2018) used deep CNN models for remote sensing image registration. Wurm, Stark, Zhu, Weigand, & Taubenböck (2019) relied on semantic segmentation to classify slum areas from aerial images.

These presented methods may differ from one another in terms of accuracies or purposes. However, the main limitation remains in how these models can be generalised to fit for multiple locations beyond the context where the models are trained and tested.

4.1.2. Seeing cities from a street-level

While it is vital to understand the overall urban systems of cities from an aerial view, seeing cities from the street-level adds more layers of information. These images can capture rapid urban changes in day-to-day life and offer more opportunities to model urban dynamics. However, capturing these rapid urban changes is a more complex task. Street-level images, taken by individuals or represented in Google's Street View API, have been used to identify a wide range of urban components from buildings to small objects such as street signs. For instance, Nguyen et al. (2018) used a CNN model to detect building types, crosswalks, and street greenness as a way to automatically quantify neighbourhood qualities.

Similarly, a range of applications based on classifying, segmenting and localising pixels from street-level images was a common approach for understanding the components of an urban scene (Chaurasia & Culurciello, 2017; Li, Jie, et al., 2017; Yang et al., 2018; Zhou et al., 2017). Scene parsing relying on semantic segmentation is a continual success of CNN models for understanding and classifying the different components of the built environment at a pixel-level (Badrinarayanan et al., 2016; Chen et al., 2016a, 2017; Lin et al., 2017; Long et al., 2015; Peng et al., 2017; Yu & Koltun, 2015; Zhao, Shi, et al., 2017; Zhao, Xiong, et al., 2017; Zhou, Zhao, et al., 2017). Relying on both street-level images and satellite images, Kang, Körner, Wang, Taubenböck, & Zhu (2018) used a deep CNN model to classify land use in satellite images by learning from building blocks of similar functions.

Quantifying the physical and non-physical appearance of cities is another area that has been intensively researched. Naik et al. (2016) quantified the physical appearance of neighbourhoods based on individuals' ranking perceptions of the urban spaces using a framework of two CNN models that are concatenated and fused to predict a score for paired street-level images, known as Streetscore-CNN. Similarly, Zhang, Xia, et al. (2018); Zhang, Zhou, et al. (2018) quantified urban spaces of street-level images labelled into six categories (Depressing, Boring, Beautiful, Safe, Lively, Wealthy) based on a crowdsourced dataset (MIT places pulse). By applying a supervised deep CNN model, they are able to predict the class for a given street view image. Liu, Silva, et al. (2017) evaluated the urban visual appearance based on two indicators of the quality of street façade and the continuity of the street walls relying on the expert ranking that is evaluated with a public survey. Moreover, Naik, Kominers, Raskar, Glaeser, & Hidalgo (2017) have used computer vision to measure the dynamics of neighbourhood characteristics from time series street view images adjoined with socioeconomic data in five US cities. As a different approach, Law, Paige, et al. (2018) used street view images to identify housing prices from urban perception relying on computer vision.

While seeing cities at street-level adds more information and gives an opportunity to understand the rapid changes that occur in an everyday urban scene in cities, the images used from Google street-view images only represent urban areas at a single weather condition, commonly clear weather, neglecting other visual and weather conditions that impact the appearance of cities. Furthermore, more research is needed on how to make best use of street level images coming from various sources, such as CCTV, dashcams or crowd sources, within and across domains.

4.2. Human interaction

Deep learning and computer vision have shown substantial progress in understanding a wide range of applications not only related to human detection but also understanding their activities and interaction with other objects (Kale & Patil, 2016; Mohamed & Ali, 2013; Zhang et al., 2017). Such approaches can assist planners and policy-makers to better understand tasks related to wellbeing and human behaviour in cities. For instance, Priya, Paul, & Singh (2015) used deep learning and computer vision to classify human actions, such as walking, running, sitting or dancing for multi-frame images. Guler, Neverova, & Kokkinos (2018) used a region-based CNN model (RCCN) to estimate the various human poses from a single image to better understand human interactions. Gkioxari, Girshick, Dollar, & He (2017) used computer vision to predict human actions over a specific target object from every day still images. This novel approach provides substantial progress in understanding human interaction with different objects. Furthermore, adjoining human pose detection with tracking, Girdhar et al. (2017) used computer vision to detect and track key human body points from videos. This could enable, for example, tackling various issues related to human safety and wellbeing in cities such as detecting when a person falls, or detecting abnormal behaviour such as crime-related actions. Indeed, a knowledge gap appears in this field of study in scaling-up deep computer vision algorithms for monitoring and detecting irregular behaviours at a city level in real-time.

4.3. Transportation and traffic

Transportation and traffic is a crucial and complex layer that merges and interacts with other layers of the city. There is a wide range of computer vision applications that aim to tackle transport modes and their common issues, such as road safety and optimisation of traffic (Buch et al., 2011; Priya et al., 2015). Subjectively, traffic surveillance and intelligent transportation systems hold the largest share of computer vision related applications in cities. Typical tasks include vehicle detection, counting, overtake detection, and traffic incident detection (Mahmud et al., 2017; Yang & Pun-Cheng, 2018). A full review of the literature on vehicle detection is beyond the scope of this article, for a comprehensive review consult Yang and Pun-Cheng (2018).

Understanding the different traffic scenarios and interactions of the different transport modes by computer vision is crucial. Bottino, Garbo, Loiacono, & Quer (2016) introduced 'Street Viewer' as a system to tackle and analyse the different scenarios of traffic behaviour from street view images. Sayed, Zaki, & Autey (2013) used computer vision to evaluate the safety measures of vehicle-bicycle conflicts. Zaki, Sayed, Tageldin, & Hussein (2013) used computer vision to analyse the conflicts among pedestrians and vehicles at a signalized intersection. Zaki & Sayed (2013) introduced a framework relying on computer vision to classify the different types of road-users.

Building on the aforementioned artificial intelligence approaches for traffic-related issues, computer vision is a core element when it comes to smart mobility and autonomous vehicles. Different applications relying on computer vision are being used to make transport modes aware of the surrounding environments either for safety indications or moving towards a self-navigation system. However, the technology of autonomous vehicles is not the focus of this research but rather the interactions of transport modes with the aforementioned layers in cities (Faisal et al., 2019).

4.4. The natural environment

The natural environment (i.e. green space, landscape, climate conditions, etc.) is a crucial layer when it comes to understanding cities. It influences our perception of the visual appearance of the built environment and also affects mobility and human interaction in cities. Different aspects related to this natural layer of cities have been tackled by computer vision. These applications vary from mapping vegetation and greenery in cities, or so-called 'Treepedia' (Cai et al., 2018), identifying plant types (Krause et al., 2018; Sun et al., 2017), to deeper understanding of the natural environment and wildlife such as detecting plant-related diseases (Mohanty et al., 2016) and understanding the patterns of social interaction among animals (Robie et al., 2017).

Deep learning and computer vision have also been used to infer the weather, climatic and air conditions in cities. Liu, Anguelov, et al. (2016) used the CNN model to identify extreme weather conditions from aerial images of climate simulations and reanalysis products. Liu, Tsow, et al. (2016) used images to analyse particle pollution for Beijing, Shanghai and Phoenix relying on region of interest selection, feature extraction and regression models. Li et al. (2019) developed a model to detect clouds from high-resolution aerial view images relying on CNNs, named multi-scale convolutional feature fusion.

While there is noticeable progress in term of methods development and accuracy enhancement among the presented papers, the common limitation remains in the lack of a single model or a framework that fuses various models to infer the different weather and environmental conditions.

4.5. Infrastructure

Cities comprise a range of infrastructure systems that represent a large portion of their economy. Inspecting these systems and detecting their deficiencies is a crucial aspect for engineers and planners in cities. The focus of this section differs from the built environment section by analysing materials and the civil engineering related issues that are not covered in the aforementioned sections.

So far, the applications of computer vision have been seen in a wide range of domains related to infrastructure and civil engineering (Gopalakrishnan, 2018; Griffiths & Boehm, 2018), most importantly in analysing defects (Feng et al., 2017). For instance, Wang, Zhao, et al. (2018) used computer vision to detect concrete crack damage. Similarly, Cha, Choi, & Büyüköztürk (2017) applied computer vision relying on deep CNN model to detect crack damage of concrete. On the other hand, Maeda, Sekimoto, Seto, Kashiyama, & Omata (2018) used computer vision to detect road damage from images that are taken from mobile devices.

5. What remains missing?

Section 3 of this paper presented the different types of computer vision algorithms that are available to researchers, and the sectors in which they have been applied were presented in Section 4. Typically, these models have been applied in a sectoral fashion to a specific problem. Comparatively little attention has been placed on how to understand the interconnections between the different layers of the city. These interconnections will eventually lead to increased capabilities of computer vision and AI to aid decision making and policy. In this section, we outline 2 under-researched areas in which computer vision has enormous potential.

5.1. Integrated models of the layers of the city

A significant challenge remains in modelling the interconnectedness and dependencies of the different layers of the city that were introduced in Fig. 2. A first step in this regard is the integration of models that have been developed for each layer in isolation. For example, there is still a knowledge gap in how to use computer vision coupled with deep learning to understand the interaction between people in cities and transport modes, or the influence of one mode on the others in terms of accessibility and safety. While the technology is there, the challenge remains in combining different models in a framework that enables them to tackle complex, multi-layered issues using the same data source, rather than just combining or fusing outputs from different data sources. On the other hand, even if the knowledge of the models is transferable among the different layers of cities, the challenges remain in finding comprehensive image data sources that cover a wide scope of tasks and functions in cities.

5.2. The scale of applying computer vision in cities

Understanding cities requires both local and global perspectives, in which scale plays a crucial role in tackling urban issues. There are different algorithms that have been used to understand, for example, individuals' actions and activities. Challenges remain in applying and scaling up such algorithms to the city level. Although there are different models, as discussed in the literature, that extract information at the city scale, the nature of the developed algorithms is still limited towards the analysis of certain area or city. The reason for this is either because of a lack of computational resources or the inability of trained models to generalise to a larger dataset at a city-level. Models often require further training and optimisation to be deployed in real-life applications. It is well known that computer vision algorithms require large sets of labelled data, which must often be manually labelled. Labels can be crowdsourced but there is often a cost involved and accuracy is difficult to guarantee. Semi- or weakly supervised learning methods are promising approaches in this regard (Guo, Chou, et al., 2018; Guo, Huang, et al., 2018).

6. Moving from prediction to decision-making to policy

After addressing the limitations of the stated models, the superior performance of modern computer vision algorithms is in little doubt. However, the extent to which model outputs can be used for automated and optimised policy and decision making remains an important research frontier. Big data, of which image data is a subset, is increasingly having an impact on decision and policy making, whether explicitly or not. Government authorities rely on algorithmic outputs to inform their decisions on a daily basis. The practical, ethical and societal implications of this are still unclear and Duarte & Álvarez (2019) note the lack of synchronicity between the potential societal impact of AI technologies and our cultural discussions around them. An option that shows promise in this direction is the concept of living labs and policy labs. These provide testbeds within which to test data driven policies, which use ICT to realise the benefits of new data sources and support collaboration with relevant stakeholders and citizens (van Veenstra & Kotterink, 2017).

Alongside other sources of big data, images and video play a particularly important role in this effort because they capture the action and interaction of humans within their environment. This provides the opportunity to understand a range of issues, such as how the structure of the built environment affects pedestrian safety, or how street lighting influences crime. These issues are inextricably linked, and urban planning and policy making must take a holistic view of them to avoid disadvantaging certain groups.

6.1. Enabling technologies

There are two enabling technologies that will be important in this area. Firstly, multi-agent reinforcement learning (MARL) will enable more realistic human agents to be simulated in more realistic urban environments. The behaviour of these agents can be learned and validated using image and video data. Such models could support or supersede traditional land use and transport planning approaches, as well as optimise the performance of urban systems such as transportation.

The second technology is GANs. It is not inconceivable that GANs, fed with images of a city, whether street view or aerial, could eventually be trained to design effective urban environments. In the same way that GANs can generate synthetic human faces that are indistinguishable from real faces (Karras et al., 2018), they could be used to plan new cities or neighbourhoods that perform like existing cities. This is certainly a long way off, but advancements in AI will enable predictions that are beyond what humans of social groups may achieve, or even conceive of (Duarte & Álvarez, 2019).

6.2. AI-embedded cameras in cities for real-time insights

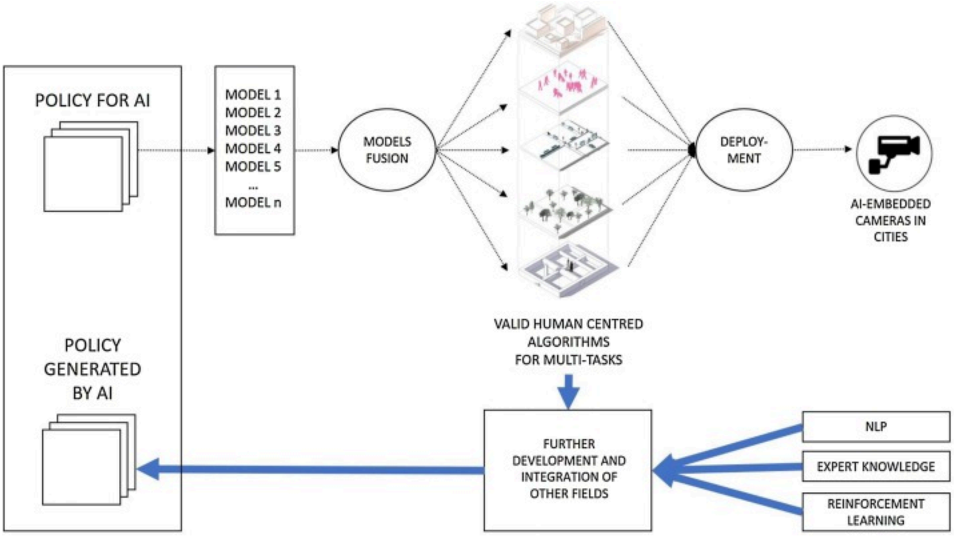
The implementation of computer vision model pipelines in (near) real-time is a crucial issue for urban analytics and Internet of Things (IoT) systems. This deployment at the edge in urban contexts can show a direct impact of the current research for developing urban theories and policies. For example, AI-embedded cameras may alert police or transport control rooms of incidents, which they can verify and respond to. This type of system should be managed in a coordinated fashion so that the needs of various authorities can be met, which requires integration of the different layers of the city. However, while this approach will enable fast decision making and response, it falls short of being a fully intelligent and automated system able to implement or generate policy.

6.3. Policy for AI and by AI

After the deployment of AI in cities based on accepted norms and ethics, their deployment in cities will also lead to the generation of adaptive urban policies by AI. AI has the potential to generate dynamic and place-based policies. However, challenges remain in the innovation and fusion of different domains of knowledge to reach this critical step where the machine not only predicts and makes decisions but generates short- and long-term plans. Most importantly, it is a mixture of the tackled deep learning and computer vision research in urban settings with Natural Language Processing (NLP) research and reinforcement learning. By merging these different knowledge domains and integrating models that are capable of addressing multiple tasks in cities, theories and more flexible place-oriented policies can be generated for cities. Nevertheless, knowledge can be transferred from one city to another.

6.4. Conceptual framework towards AI generated policy and decision making

Fig. 3 shows a conceptual framework and a recommended process for achieving the two crucial steps outlined in Sections 6.2 and 6.3, and how they can be reached from the current perspective of the deep computer vision research that is highlighted in this review. It shows the overall system for policy-makers and developers showing the important aspect of this process and the domains that are still under-developed and require further integration with urban analytics research.



Download : Download high-res image (355KB)

Download : Download full-size image

Fig. 3. Conceptual framework for AI-generated urban policies. The figure is created by the authors.

Currently we are at the stage where policy for AI is being developed to mitigate the risks of reliance on the technology to make decisions. However, we envisage a future where the integration of the layers of the city through AI enables understanding of urban processes that is not possible by viewing them in isolation, leading to AI generated policies, as stated in Section 6.3.

7. Conclusions

Understanding cities has been a profound interest for many scholars across a wide range of disciplines. Modelling the different urban systems of cities is a longevity purpose for many urban and transport planners. While cities are complex by nature and classical urban modelling may not capture the actual complexities of urban systems, computer vision shows progress in tackling a variety of complex physical and non-physical visual tasks. In this article, we provide a review of deep learning and computer vision and its application so far in understanding cities. The article highlights the different types of algorithms of computer vision and their application to cities and their multifaced issues. It aimed to show the nuances of the variations of these algorithms within the same task. It also aimed to show what has been done so far to understand cities by machine vision and what remains missing for future research work within this domain.

We attempt to highlight the potential role of computer vision in understanding the interactions between the built environment, people and transportation in order to tackle the complexity and nonlinearity of many urban and transport issues for better policy-making and planning safer cities. We also highlight the current limitations that require further work to reach an integrated computer vision-based urban models that capable of making automatic decisions. We also explain how integrated computer vision models, with other knowledge domains, can be embedded in cities, not only for prediction and decision-making but also to generate flexible place-oriented policies.

Acknowledgement

This research outcome is a part of a PhD study for the first author at University College London (UCL). It was supported by funds from UCL Overseas Research Scholarship (ORS) and the Road Safety Trust (RST 38_03_2017).

Recommended articles

References

Amirkolaee and Arefi, 2019 H.A. Amirkolaee, H. Arefi

Height estimation from single aerial images using a deep convolutional encoder-decoder network

ISPRS Journal of Photogrammetry and Remote Sensing, 149 (2019), pp. 50-66, 10.1016/j.isprsjprs.2019.01.013 ↗

View PDF View article View in Scopus ↗ Google Scholar ↗

Audebert et al., 2018 N. Audebert, B. Le Saux, S. Lefèvre

Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks

ISPRS Journal of Photogrammetry and Remote Sensing, 140 (2018), pp. 20-32, 10.1016/j.isprsjprs.2017.11.011 ↗

View PDF View article View in Scopus ↗ Google Scholar ↗

Badrinarayanan et al., 2016 V. Badrinarayanan, A. Kendall, R. Cipolla

SegNet: A deep convolutional encoder-decoder architecture for image segmentation

arXiv:1511.00561v3 [cs.CV]

(2016)

Google Scholar ↗

Batty, 2008 M. Batty

The size, scale, and shape of cities

Science, 319 (5864) (2008), pp. 769-771, 10.1126/science.1151419 ↗

View in Scopus ↗ Google Scholar ↗

Batty, 2009 M. Batty

Urban modeling

International encyclopedia of human geography, Elsevier, Oxford, UK (2009), pp. 51-58

View PDF View article View in Scopus ↗ Google Scholar ↗

Batty and Longley, 1994 M. Batty, P. Longley

Fractal cities: A geometry of form and function

Academic Press, New York (1994)

Google Scholar ↗

Batty and Xie, 1996 M. Batty, Y. Xie

Preliminary evidence for a theory of the fractal city

Environment & Planning A, 28 (10) (1996), pp. 1745-1762, 10.1068/a281745 ↗

View in Scopus ↗ Google Scholar ↗

Batty et al., 1997 M. Batty, H. Couclelis, M. Eichen

Urban systems as cellular automata

SAGE Publications Sage UK, London, England (1997)

Google Scholar ↗

Batty, 2005 M. Batty

Agents, cells, and cities: New representational models for simulating multiscale urban dynamics

Environment & Planning A, 37 (8) (2005), pp. 1373-1394, 10.1068/a3784 ↗

View in Scopus ↗ Google Scholar ↗

Batty, 2019 M. Batty

Urban analytics defined

Environment and Planning B Urban Analytics and City Science, 46 (3) (2019), pp. 403-405, 10.1177/2399808319839494 ↗

View in Scopus ↗ Google Scholar ↗

Becattini et al., 2017 F. Becattini, T. Uricchio, L. Seidenari, A. Del Bimbo, L. Ballan

Am I done? Predicting action progress in videos

ArXiv:1705.01781 [Cs]. Retrieved from

(2017)

http://arxiv.org/abs/1705.01781 ↗

Google Scholar ↗

Bettencourt, 2013 L. Bettencourt

The origins of scaling in cities

Science, 340 (6139) (2013), pp. 1438-1441

CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Bilen et al., 2016 H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, S. Gould

Dynamic image networks for action recognition

2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016), pp. 3034-3042, 10.1109/CVPR.2016.331 ↗

View in Scopus ↗ Google Scholar ↗

Bottino et al., 2016 A. Bottino, A. Garbo, C. Loiacono, S. Quer

Street viewer: An autonomous vision based traffic tracking system

Sensors, 16 (6) (2016), p. 813, 10.3390/s16060813 ↗

View in Scopus ↗ Google Scholar ↗

Brock et al., 2018 A. Brock, J. Donahue, K. Simonyan

Large scale GAN training for high fidelity natural image synthesis

ArXiv:1809.11096 [Cs, Stat]. Retrieved from

(2018)

http://arxiv.org/abs/1809.11096 ↗

Google Scholar ↗

Buch et al., 2011 N. Buch, S.A. Velastin, J. Orwell

A review of computer vision techniques for the analysis of urban traffic

IEEE Transactions on Intelligent Transportation Systems, 12 (3) (2011), pp. 920-939, 10.1109/TITS.2011.2119372 ↗

View in Scopus ↗ Google Scholar ↗

Buch et al., 2017 S. Buch, V. Escorcia, C. Shen, B. Ghanem, J.C. Niebles

SST: Single-stream temporal action proposals

2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017), pp. 6373-6382, 10.1109/CVPR.2017.675 ↗

View in Scopus ↗ Google Scholar ↗

Cai et al., 2018 B.Y. Cai, X. Li, I. Seiferling, C. Ratti

Treepedia 2.0: Applying deep learning for large-scale quantification of urban tree cover

ArXiv:1808.04754 [Cs]. Retrieved from

(2018)

http://arxiv.org/abs/1808.04754 ↗

Google Scholar ↗

Calder et al., 2018 M. Calder, C. Craig, D. Culley, R. de Cani, C.A. Donnelly, R. Douglas, A. Wilson

Computational modelling for decision-making: Where, why, what, who and how

Royal Society Open Science, 5 (6) (2018), Article 172096, 10.1098/rsos.172096 ↗

View in Scopus ↗ Google Scholar ↗

Cao et al., 2017 Y. Cao, Z. Wu, C. Shen

Estimating depth from monocular images as classification using deep fully convolutional residual networks

ArXiv:1605.02305 [Cs]. Retrieved from

(2017)

http://arxiv.org/abs/1605.02305 ↗

Google Scholar ↗

Cao et al., 2016 Z. Cao, T. Simon, S.-E. Wei, Y. Sheikh

Realtime multi-person 2D pose estimation using part affinity fields

ArXiv:1611.08050 [Cs]. Retrieved from

(2016)

http://arxiv.org/abs/1611.08050 ↗

Google Scholar ↗

Caron et al., 2018 M. Caron, P. Bojanowski, A. Joulin, M. Douze

Deep clustering for unsupervised learning of visual features

(2018), p. 29

CrossRef ↗ Google Scholar ↗

Cha et al., 2017 Y.-J. Cha, W. Choi, O. Büyüköztürk

Deep learning-based crack damage detection using convolutional neural networks: Deep learning-based crack damage detection using CNNs
Computer-Aided Civil and Infrastructure Engineering, 32 (5) (2017), pp. 361-378, 10.1111/mice.12263 ↗
View in Scopus ↗ Google Scholar ↗

Chao et al., 2018 Y.-W. Chao, S. Vijayanarasimhan, B. Seybold, D.A. Ross, J. Deng, R. Sukthankar
Rethinking the faster R-CNN architecture for temporal action localization
ArXiv:1804.07667 [Cs]. Retrieved from
(2018)
http://arxiv.org/abs/1804.07667 ↗
Google Scholar ↗

Chaurasia and Culurciello, 2017 A. Chaurasia, E. Culurciello
LinkNet: Exploiting encoder representations for efficient semantic segmentation
2017 IEEE Visual Communications and Image Processing (VCIP) (2017), pp. 1-4, 10.1109/VCIP.2017.8305148 ↗
Google Scholar ↗

Chen, Dowman, et al., 2016 J. Chen, I. Dowman, S. Li, Z. Li, M. Madden, J. Mills, C. Heipke
Information from imagery: ISPRS scientific vision and research agenda
ISPRS Journal of Photogrammetry and Remote Sensing, 115 (2016), pp. 3-21, 10.1016/j.isprsjprs.2015.09.008 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Chen, Yang, et al., 2016 Y. Chen, X. Yang, B. Zhong, S. Pan, D. Chen, H. Zhang
CNNTracker: Online discriminative object tracking via deep convolutional neural network
Applied Soft Computing, 38 (2016), pp. 1088-1098, 10.1016/j.asoc.2015.06.048 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Chen et al., 2016a L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. Yuille
Semantic Image segmentation with deep convolutional nets and fully connected CRFS
(2016)
Google Scholar ↗

Chen et al., 2016b L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille
DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs
ArXiv:1606.00915 [Cs]. Retrieved from
(2016)
http://arxiv.org/abs/1606.00915 ↗
Google Scholar ↗

Chen et al., 2017 L.-C. Chen, G. Papandreou, F. Schroff, H. Adam
Rethinking atrous convolution for semantic image segmentation
ArXiv Preprint ArXiv:1706.05587
(2017)
Google Scholar ↗

Chen and Corso, 2015 W. Chen, J.J. Corso
Action detection by implicit intentional motion clustering
2015 IEEE International Conference on Computer Vision (ICCV) (2015), pp. 3298-3306, 10.1109/ICCV.2015.377 ↗
Google Scholar ↗

Chew, Amer, et al., 2018 R.F. Chew, S. Amer, K. Jones, J. Unangst, J. Cajka, J. Allpress, M. Bruhn
Residential scene classification for gridded population sampling in developing countries using deep convolutional neural networks on satellite imagery
International Journal of Health Geographics, 17 (1) (2018), 10.1186/s12942-018-0132-1 ↗
Google Scholar ↗

Chew, Jones, et al., 2018 R. Chew, K. Jones, J. Unangst, J. Cajka, J. Allpress, S. Amer, K. Krotki
Toward model-generated household listing in low- and middle-income countries using deep learning
ISPRS International Journal of Geo-information, 7 (11) (2018), p. 448, 10.3390/ijgi7110448 ↗
View in Scopus ↗ Google Scholar ↗

Chu et al., 2017 Q. Chu, W. Ouyang, H. Li, X. Wang, B. Liu, N. Yu

Online multi-object tracking using CNN-based single object tracker with spatial-temporal attention mechanism
ArXiv:1708.02843 [Cs]. Retrieved from
(2017)
http://arxiv.org/abs/1708.02843 ↗
Google Scholar ↗

Cordts et al., 2016 M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, B. Schiele
The cityscapes dataset for semantic urban scene understanding
Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016), pp. 3213-3223
View in Scopus ↗ Google Scholar ↗

Danelljan et al., 2016 M. Danelljan, G. Bhat, F.S. Khan, M. Felsberg
ECO: Efficient convolution operators for tracking
ArXiv:1611.09224 [Cs]. Retrieved from
(2016)
http://arxiv.org/abs/1611.09224 ↗
Google Scholar ↗

Danelljan et al., 2015 M. Danelljan, G. Hager, F.S. Khan, M. Felsberg
Convolutional features for correlation filter based visual tracking
2015 IEEE International Conference on Computer Vision Workshop (ICCVW) (2015), pp. 621-629, 10.1109/ICCVW.2015.84 ↗
View in Scopus ↗ Google Scholar ↗

De Nadai et al., 2016 M. De Nadai, R.L. Vieriu, G. Zen, S. Dragicevic, N. Naik, M. Caraviello, ..., B. Lepri
Are safer looking neighborhoods more lively?: A multimodal investigation into Urban life
Proceedings of the 2016 ACM on Multimedia Conference - MM' 16 (2016), pp. 1127-1135, 10.1145/2964284.2964312 ↗
View in Scopus ↗ Google Scholar ↗

Demir et al., 2018 I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, R. Raskar
DeepGlobe 2018: A challenge to parse the Earth through satellite images
(2018), p. 10
View in Scopus ↗ Google Scholar ↗

Diba et al., 2017 A. Diba, M. Fayyaz, V. Sharma, A.H. Karami, M.M. Arzani, R. Yousefzadeh, L.V. Gool
Temporal 3D ConvNets using temporal transition layer
(2017), p. 5
Google Scholar ↗

Duarte and Álvarez, 2019 F. Duarte, R. Álvarez
The data politics of the urban age
Palgrave Communications, 5 (1) (2019), pp. 1-7, 10.1057/s41599-019-0264-3 ↗
Google Scholar ↗

Dubey et al., 2016 A. Dubey, N. Naik, D. Parikh, R. Raskar, C.A. Hidalgo
Deep learning the city: Quantifying urban perception at a global scale
European Conference on Computer Vision (2016), pp. 196-212
Springer
CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Elhoseiny et al., 2015 M. Elhoseiny, S. Huang, A. Elgammal
Weather classification with deep convolutional neural networks. 2015
IEEE International Conference on Image Processing (ICIP) (2015), pp. 3349-3353, 10.1109/ICIP.2015.7351424 ↗
View in Scopus ↗ Google Scholar ↗

El-Nouby and Taylor, 2018 A. El-Nouby, G.W. Taylor
Real-time end-to-End action detection with two-stream networks
ArXiv:1802.08362 [Cs]. Retrieved from
(2018)
http://arxiv.org/abs/1802.08362 ↗
Google Scholar ↗

Escorcia et al., 2016 Escorcia, V., Caba Heilbron, F., Niebles, J. C., & Ghanem, B. (2016). DAPs: Deep Action Proposals for Action Understanding. In B. Leibe, J. Matas, N. Sebe, & M. Welling, Computer Vision – ECCV 2016 (Vol. 9907, pp. 768–784). https://doi.org/10.1007/978-3-319-46487-9_47.

Google Scholar ↗

Eslami et al., 2018 S.M.A. Eslami, D. Jimenez Rezende, F. Besse, F. Viola, A.S. Morcos, M. Garnelo, ..., D. Hassabis
Neural scene representation and rendering
Science, 360 (6394) (2018), pp. 1204-1210, 10.1126/science.aar6170 ↗

Google Scholar ↗

Faisal et al., 2019 A. Faisal, T. Yigitcanlar, M. Kamruzzaman, G. Currie
Understanding autonomous vehicles: A systematic literature review on capability, impact, planning and policy
Journal of Transport and Land Use, 12 (1) (2019), 10.5198/jtlu.2019.1405 ↗

Google Scholar ↗

Fang et al., 2016 H.-S. Fang, S. Xie, Y.-W. Tai, C. Lu
RMPE: Regional multi-person pose estimation
ArXiv:1612.00137 [Cs]. Retrieved from
(2016)
http://arxiv.org/abs/1612.00137 ↗

Google Scholar ↗

Feng et al., 2017 C. Feng, M.-Y. Liu, C.-C. Kao, T.-Y. Lee
Deep active learning for civil infrastructure defect detection and classification
Computing in Civil Engineering, 2017 (2017), pp. 298-306, 10.1061/9780784480823.036 ↗

View in Scopus ↗ Google Scholar ↗

Frankhauser, 1998 P. Frankhauser
The fractal approach. A new tool for the spatial analysis of urban agglomerations
Population, 10 (1) (1998), pp. 205-240

CrossRef ↗ Google Scholar ↗

Gemert et al., 2015 J. Gemert, C. van, M. Jain, E. Gati, C.G.M. Snoek
APT: Action localization proposals from dense trajectories
Procedings of the British Machine Vision Conference 2015 (2015), 10.5244/C.29.177 ↗
177.1-177.12

Google Scholar ↗

Girdhar et al., 2017 R. Girdhar, G. Gkioxari, L. Torresani, M. Paluri, D. Tran
Detect-and-Track: Efficient pose estimation in videos
(2017), p. 10

Google Scholar ↗

Girdhar and Ramanan, 2017 R. Girdhar, D. Ramanan
Attentional pooling for action recognition
ArXiv:1711.01467 [Cs]. Retrieved from
(2017)
http://arxiv.org/abs/1711.01467 ↗

Google Scholar ↗

Girshick et al., 2014 R. Girshick, J. Donahue, T. Darrell, J. Malik
Rich feature hierarchies for accurate object detection and semantic segmentation
Retrieved from
(2014)
https://arxiv.org/pdf/1311.2524.pdf ↗

Google Scholar ↗

Gkioxari et al., 2017 G. Gkioxari, R. Girshick, P. Dollar, K. He
Detecting and recognizing human-object interactions
(2017), p. 9

Google Scholar ↗

Goodfellow, 2016 I. Goodfellow

NIPS 2016 tutorial: Generative adversarial networks

ArXiv:1701.00160 [Cs]. Retrieved from

(2016)

http://arxiv.org/abs/1701.00160 ↗

Google Scholar ↗

Goodfellow et al., 2017 I. Goodfellow, Y. Bengio, A. Courville

Deep learning

The MIT Press, Cambridge, Massachusetts (2017)

Google Scholar ↗

Goodfellow et al., 2014 I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, ..., Y. Bengio

Generative adversarial networks

ArXiv:1406.2661 [Cs, Stat]. Retrieved from

(2014)

http://arxiv.org/abs/1406.2661 ↗

Google Scholar ↗

Gopalakrishnan, 2018 K. Gopalakrishnan

Deep Learning in Data-Driven Pavement Image Analysis and Automated Distress Detection: A Review

Data, 3 (3) (2018), p. 28, 10.3390/data3030028 ↗

View in Scopus ↗ Google Scholar ↗

Griffiths and Boehm, 2018 D. Griffiths, J. Boehm

RAPID OBJECT DETECTION SYSTEMS, UTILISING DEEP LEARNING AND UNMANNED AERIAL SYSTEMS (UAS) FOR CIVIL ENGINEERING APPLICATIONS

ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII–2 (2018), pp. 391-398, 10.5194/isprs-archives-XLII-2-391-2018 ↗

View in Scopus ↗ Google Scholar ↗

Guerra et al., 2018 J.C.V. Guerra, Z. Khanam, S. Ehsan, R. Stolkin, K. McDonald-Maier

Weather Classification: A new multi-class dataset, data augmentation approach and comprehensive evaluations of Convolutional Neural Networks

ArXiv:1808.00588 [Cs]. Retrieved from

(2018)

http://arxiv.org/abs/1808.00588 ↗

Google Scholar ↗

Guler et al., 2018 R.A. Guler, N. Neverova, I. Kokkinos

DensePose: Dense human pose estimation in the wild

(2018), p. 10

Google Scholar ↗

Guo, Chou, et al., 2018 M. Guo, E. Chou, D.-A. Huang, S. Song, S. Yeung, L. Fei-Fei

Neural graph matching networks for fewshot 3D action recognition

(2018), p. 17

CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Guo, Huang, et al., 2018 S. Guo, W. Huang, H. Zhang, C. Zhuang, D. Dong, M.R. Scott, D. Huang

CurriculumNet: Weakly supervised learning from large-scale web images

ArXiv:1808.01097 [Cs]. Retrieved from

(2018)

http://arxiv.org/abs/1808.01097 ↗

Google Scholar ↗

Guo et al., 2016 Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, M.S. Lew

Deep learning for visual understanding: A review

Neurocomputing, 187 (2016), pp. 27-48, 10.1016/j.neucom.2015.09.116 ↗

View PDF View article View in Scopus ↗ Google Scholar ↗

He et al., 2019 H. He, D. Yang, S. Wang, S. Wang, Y. Li
Road extraction by using atrous spatial pyramid pooling integrated encoder-decoder network and structural similarity loss
Remote Sensing, 11 (9) (2019), p. 1015, 10.3390/rs11091015
View in Scopus Google Scholar

He et al., 2015 K. He, X. Zhang, S. Ren, J. Sun
Deep residual learning for image recognition
ArXiv:1512.03385v1. Retrieved from
(2015)
https://arxiv.org/pdf/1512.03385.pdf
Google Scholar

He et al., 2018 L. He, G. Wang, Z. Hu
Learning depth from single images with deep neural network embedding focal length
IEEE Transactions on Image Processing, 27 (9) (2018), pp. 4676-4689, 10.1109/TIP.2018.2832296
View in Scopus Google Scholar

Helbich et al., 2019 M. Helbich, Y. Yao, Y. Liu, J. Zhang, P. Liu, R. Wang
Using deep learning to examine street view green and blue spaces and their associations with geriatric depression in Beijing
China. Environment International, 126 (2019), pp. 107-117, 10.1016/j.envint.2019.02.013
View PDF View article View in Scopus Google Scholar

Held et al., 2016 D. Held, S. Thrun, S. Savarese
Learning to track at 100 FPS with deep regression networks
(2016)
ArXiv:1604.01802 [Cs]. Retrieved from http://arxiv.org/abs/1604.01802
Google Scholar

Heppenstall et al., 2012 Heppenstall, A. J., Crooks, A. T., See, L. M., & Batty, M. (Eds.). (2012). Agent-Based Models of Geographical Systems. https://doi.org/10.1007/978-90-481-8927-4.
Google Scholar

Hester et al., 2017 Hester, T., Vecerik, M., Pietquin, O., Lanctot, M., Schaul, T., Piot, B., Gruslys, A. (2017). Deep Q-learning from Demonstrations. ArXiv:1704.03732 [Cs]. Retrieved from http://arxiv.org/abs/1704.03732.
Google Scholar

Hong et al., 2019 S.-J. Hong, Y. Han, S.-Y. Kim, A.-Y. Lee, G. Kim
Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery
Sensors, 19 (7) (2019), p. 1651, 10.3390/s19071651
View in Scopus Google Scholar

Hou et al., 2017 R. Hou, C. Chen, M. Shah
Tube convolutional neural network (T-CNN) for action detection in videos
2017 IEEE International Conference on Computer Vision (ICCV) (2017), pp. 5823-5832, 10.1109/ICCV.2017.620
View in Scopus Google Scholar

Huang et al., 2017 Huang, G., Liu, Z., Weinberger, K. Q., & van der Maaten, L. (2017). Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1, 3.
Google Scholar

Ibrahim et al., 2019 M.R. Ibrahim, J. Haworth, T. Cheng
URBAN-i: From urban scenes to mapping slums, transport modes, and pedestrians in cities using deep learning and computer vision
Environment and Planning B Urban Analytics and City Science (2019), 10.1177/2399808319846517
239980831984651
Google Scholar

Insafutdinov, Andriluka, et al., 2016 E. Insafutdinov, M. Andriluka, L. Pishchulin, S. Tang, E. Levinkov, B. Andres, B. Schiele
ArtTrack: Articulated multi-person tracking in the wild
ArXiv:1612.01465 [Cs]. Retrieved from

(2016)

http://arxiv.org/abs/1612.01465 ↗

Google Scholar ↗

Insafutdinov, Pishchulin, et al., 2016 E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, B. Schiele

DeeperCut: A deeper, stronger, and faster multi-person pose estimation model

ArXiv:1605.03170 [Cs]. Retrieved from

(2016)

http://arxiv.org/abs/1605.03170 ↗

Google Scholar ↗

Isalgue et al., 2007 A. Isalgue, H. Coch, R. Serra

Scaling laws and the modern city

Physica A Statistical Mechanics and Its Applications, 382 (2) (2007), pp. 643-649, 10.1016/j.physa.2007.04.019 ↗

View PDF View article View in Scopus ↗ Google Scholar ↗

Isola et al., 2016 P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros

Image-to-Image translation with conditional adversarial networks

ArXiv:1611.07004 [Cs]. Retrieved from

(2016)

http://arxiv.org/abs/1611.07004 ↗

Google Scholar ↗

Jégou et al., 2016 S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, Y. Bengio

The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation

ArXiv:1611.09326 [Cs]. Retrieved from

(2016)

http://arxiv.org/abs/1611.09326 ↗

Google Scholar ↗

Jiang et al., 2018 C. Jiang, J. Xiao, Y. Xie, T. Tillo, K. Huang

Siamese network ensemble for visual tracking

Neurocomputing, 275 (2018), pp. 2892-2903, 10.1016/j.neucom.2017.10.043 ↗

View PDF View article View in Scopus ↗ Google Scholar ↗

Kale and Patil, 2016 G.V. Kale, V.H. Patil

A study of vision based human motion recognition and analysis

International Journal of Ambient Computing and Intelligence, 7 (2) (2016), p. 18

Google Scholar ↗

Kang et al., 2018 J. Kang, M. Körner, Y. Wang, H. Taubenböck, X.X. Zhu

Building instance classification using street view images

ISPRS Journal of Photogrammetry and Remote Sensing, 145 (2018), pp. 44-59, 10.1016/j.isprsjprs.2018.02.006 ↗

View PDF View article View in Scopus ↗ Google Scholar ↗

Kang et al., 2016 K. Kang, W. Ouyang, H. Li, X. Wang

Object detection from video tubelets with convolutional neural networks

2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016), pp. 817-825, 10.1109/CVPR.2016.95 ↗

View in Scopus ↗ Google Scholar ↗

Karras et al., 2018 T. Karras, S. Laine, T. Aila

A style-based generator architecture for generative adversarial networks

ArXiv:1812.04948 [Cs, Stat]. Retrieved from

(2018)

http://arxiv.org/abs/1812.04948 ↗

Google Scholar ↗

Kocabas et al., 2018 M. Kocabas, S. Karagoz, E. Akbas

MultiPoseNet: Fast multi-person pose estimation using pose residual network

ArXiv:1807.04067 [Cs]. Retrieved from

(2018)

http://arxiv.org/abs/1807.04067 ↗

Google Scholar

Krause et al., 2018 J. Krause, G. Sugita, K. Baek, L. Lim
WTPlant(what's that plant?): A deep learning system for identifying plants in natural images
Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval - ICMR' 18 (2018), pp. 517-520, 10.1145/3206025.3206089
View in Scopus Google Scholar

Krizhevsky et al., 2012 A. Krizhevsky, I. Sutskever, G.E. Hinton
Imagenet classification with deep convolutional neural networks
Proceeding NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems, 1, Lake Tahoe, Nevada: Curran Associates Inc., USA ©2012 (2012), pp. 1097-1105
Google Scholar

Kuo, 2016 C.-C.J. Kuo
Understanding convolutional neural networks with a mathematical model
Journal of Visual Communication and Image Representation, 41 (2016), pp. 406-413
View PDF View article View in Scopus Google Scholar

Law, Paige, et al., 2018 S. Law, B. Paige, C. Russell
Take a look around: Using street view and satellite images to estimate house prices
ArXiv:1807.07155 [Cs, Econ]
(2018)
Retrieved from
http://arxiv.org/abs/1807.07155
Google Scholar

Law, Seresinhe, et al., 2018 S. Law, C.I. Seresinhe, Y. Shen, M. Gutierrez-Roig
Street-Frontage-Net: Urban image classification using deep convolutional neural networks
International Journal of Geographical Information Science (2018), pp. 1-27, 10.1080/13658816.2018.1555832
Google Scholar

LeCun et al., 2015 Y. LeCun, Y. Bengio, G. Hinton
Deep learning
Nature, 521 (7553) (2015), pp. 436-444, 10.1038/nature14539
View in Scopus Google Scholar

Li et al., 2018 P. Li, D. Wang, L. Wang, H. Lu
Deep visual tracking: Review and experimental comparison
Pattern Recognition, 76 (2018), pp. 323-338, 10.1016/j.patcog.2017.11.007
View PDF View article View in Scopus Google Scholar

Li, Jie, et al., 2017 X. Li, Z. Jie, W. Wang, C. Liu, J. Yang, X. Shen, J. Feng
FoveaNet: Perspective-aware urban scene parsing
ArXiv:1708.02421 [Cs]. Retrieved from
(2017)
http://arxiv.org/abs/1708.02421
Google Scholar

Li, Wang, et al., 2017 X. Li, Z.J.W. Wang, C.L.J. Yang, X.S.Z.L.Q. Chen, S. Yan, J. Feng
FoveaNet: Perspective-aware urban scene parsing
ArXiv Preprint ArXiv:1708.02421
(2017)
Google Scholar

Li et al., 2016 Y. Li, C. Lan, J. Xing, W. Zeng, C. Yuan, J. Liu
Online human action detection using joint classification-regression recurrent neural networks
ArXiv:1604.05633 [Cs]. Retrieved from
(2016)
http://arxiv.org/abs/1604.05633
Google Scholar

Li et al., 2019 Z. Li, H. Shen, Q. Cheng, Y. Liu, S. You, Z. He

Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors
ISPRS Journal of Photogrammetry and Remote Sensing, 150 (2019), pp. 197-212, 10.1016/j.isprsjprs.2019.02.017 ↗

View PDF View article Google Scholar ↗

Lin et al., 2017 G. Lin, A. Milan, C. Shen, I. Reid
Refinenet: Multi-path refinement networks for high-resolution semantic segmentation
IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
Google Scholar ↗

Lin et al., 2018 T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár
Focal loss for dense object detection
ArXiv:1708.02002 [Cs]. Retrieved from
(2018)
http://arxiv.org/abs/1708.02002 ↗
Google Scholar ↗

Lin et al., 2014 T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, ..., P. Dollár
Microsoft COCO: Common objects in context
ArXiv:1405.0312 [Cs]. Retrieved from
(2014)
http://arxiv.org/abs/1405.0312 ↗
Google Scholar ↗

Liu, Tsow, et al., 2016 C. Liu, F. Tsow, Y. Zou, N. Tao
Particle pollution estimation based on image analysis
PloS One, 11 (2) (2016), Article e0145955, 10.1371/journal.pone.0145955 ↗
View in Scopus ↗ Google Scholar ↗

Liu, Silva, et al., 2017 L. Liu, E.A. Silva, C. Wu, H. Wang
A machine learning-based method for the large-scale evaluation of the qualities of the urban environment
Computers, Environment and Urban Systems, 65 (2017), pp. 113-125, 10.1016/j.compenvurbsys.2017.06.003 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Liu, Anguelov, et al., 2016 W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg
Ssd: Single shot multibox detector
Computer Vision - ECCV : European Conference on Computer Vision : Proceedings European Conference on Computer Vision (2016), pp. 21-37
Springer
View in Scopus ↗ Google Scholar ↗

Liu, Yang, et al., 2017 W. Liu, Y. Yang, L. Wei, School of Automation, China University of Geosciences
Weather recognition of street scene based on sparse deep neural networks
Journal of Advanced Computational Intelligence and Intelligent Informatics, 21 (3) (2017), pp. 403-408, 10.20965/jaciii.2017.p0403 ↗
View in Scopus ↗ Google Scholar ↗

Liu, Racah, et al., 2016 Y. Liu, E. Racah, J. Prabhat Correa, A. Khosrowshahi, D. Lavers, ..., W. Collins
Application of deep convolutional neural networks for detecting extreme weather in climate datasets
ArXiv:1605.01156 [Cs]. Retrieved from
(2016)
http://arxiv.org/abs/1605.01156 ↗
Google Scholar ↗

Long et al., 2015 J. Long, E. Shelhamer, T. Darrell
Fully convolutional networks for semantic segmentation
arXiv:1411.4038v2 [cs.CV]
(2015)
Google Scholar ↗

Maeda et al., 2018 H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, H. Omata
Road damage detection using deep neural networks with images captured through a smartphone
ArXiv:1801.09454 [Cs]. Retrieved from
(2018)
http://arxiv.org/abs/1801.09454 ↗

Google Scholar ↗

Mahmud et al., 2017 S.M.S. Mahmud, L. Ferreira, M.S. Hoque, A. Tavassoli
Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs
IATSS Research, 41 (4) (2017), pp. 153-163, 10.1016/j.iatssr.2017.02.001 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Manen et al., 2017 S. Manen, M. Gygli, D. Dai, L.V. Gool
PathTrack: Fast trajectory annotation with path supervision
2017 IEEE International Conference on Computer Vision (ICCV) (2017), pp. 290-299, 10.1109/ICCV.2017.40 ↗
View in Scopus ↗ Google Scholar ↗

Marcos et al., 2018 D. Marcos, M. Volpi, B. Kellenberger, D. Tuia
Land cover mapping at very high resolution with rotation equivariant CNNs: Towards small yet accurate models
ISPRS Journal of Photogrammetry and Remote Sensing, 145 (2018), pp. 96-107, 10.1016/j.isprsjprs.2018.01.021 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Mettes et al., 2016 P. Mettes, J.C. van Gemert, C.G.M. Snoek
Spot on: Action localization from pointly-supervised proposals
B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Computer vision – ECCV 2016, Vol. 9909 (2016), 10.1007/978-3-319-46454-1_27 ↗
437–453
Google Scholar ↗

Mirowski et al., 2018 P. Mirowski, M.K. Grimes, M. Malinowski, K.M. Hermann, K. Anderson, D. Teplyashin, ..., R. Hadsell
Learning to navigate in cities without a map
ArXiv:1804.00168 [Cs]. Retrieved from
(2018)
http://arxiv.org/abs/1804.00168 ↗
Google Scholar ↗

Mnih et al., 2016 V. Mnih, A.P. Badia, M. Mirza, A. Graves, T. Harley, T.P. Lillicrap, ..., K. Kavukcuoglu
Asynchronous methods for deep reinforcement learning
(2016), p. 10
Google Scholar ↗

Mnih et al., 2013 V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller
Playing atari with deep reinforcement learning
ArXiv:1312.5602 [Cs]. Retrieved from
(2013)
http://arxiv.org/abs/1312.5602 ↗
Google Scholar ↗

Mohamed and Ali, 2013 A.N. Mohamed, M.M. Ali
Human motion analysis, recognition and understanding in Computer Vision: A REVIEW
Journal of Engineering Sciences, 41 (5) (2013), p. 19
Google Scholar ↗

Mohanty et al., 2016 S.P. Mohanty, D.P. Hughes, M. Salathé
Using deep learning for image-based plant disease detection
Frontiers in Plant Science, 7 (2016), 10.3389/fpls.2016.01419 ↗
Google Scholar ↗

Murcio et al., 2015 R. Murcio, A.P. Masucci, E. Arcaute, M. Batty
Multifractal to monofractal evolution of the London street network
Physical Review E, 92 (6) (2015), 10.1103/PhysRevE.92.062130 ↗
Google Scholar ↗

Naik et al., 2017 N. Naik, S.D. Kominers, R. Raskar, E.L. Glaeser, C.A. Hidalgo
Computer vision uncovers predictors of physical urban change
Proceedings of the National Academy of Sciences, 114 (29) (2017), pp. 7571-7576, 10.1073/pnas.1619003114 ↗
View in Scopus ↗ Google Scholar ↗

Naik et al., 2014 N. Naik, J. Philipoom, R. Raskar, C. Hidalgo
Streetscore-predicting the perceived safety of one million streetscapes
Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2014), pp. 779-785
Google Scholar ↗

Naik et al., 2016 N. Naik, R. Raskar, C.A. Hidalgo
Cities are physical too: Using computer vision to measure the quality and impact of urban appearance
The American Economic Review, 106 (5) (2016), pp. 128-132, 10.1257/aer.p20161030 ↗
View in Scopus ↗ Google Scholar ↗

Narazaki et al., 2017 Y. Narazaki, V. Hoskere, T.A. Hoang, B.F.S. Jr
Vision-based automated bridge component recognition integrated with high-level scene understanding
(2017), p. 10
Google Scholar ↗

Nguyen et al., 2018 Q.C. Nguyen, M. Sajjadi, M. McCullough, M. Pham, T.T. Nguyen, W. Yu, ..., T. Tasdizen
Neighbourhood looking glass: 360° automated characterisation of the built environment for neighbourhood effects research
Journal of Epidemiology and Community Health, 72 (3) (2018), pp. 260-266, 10.1136/jech-2017-209456 ↗
View in Scopus ↗ Google Scholar ↗

Oliva and Torralba, 2006 A. Oliva, A. Torralba
Chapter 2 Building the gist of a scene: The role of global image features in recognition
Progress in Brain Research, 155 (2006), pp. 23-36, 10.1016/S0079-6123(06)55002-2 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Paganini et al., 2018 M. Paganini, L. de Oliveira, B. Nachman
CaloGAN: Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks
Physical Review D, 97 (1) (2018), 10.1103/PhysRevD.97.014021 ↗
Google Scholar ↗

Papandreou et al., 2017 G. Papandreou, T. Zhu, N. Kanazawa, A. Toshev, J. Tompson, C. Bregler, K. Murphy
Towards accurate multi-person pose estimation in the wild
ArXiv:1701.01779 [Cs]. Retrieved from
(2017)
http://arxiv.org/abs/1701.01779 ↗
Google Scholar ↗

Peng et al., 2017 C. Peng, X. Zhang, G. Yu, G. Luo, J. Sun
Large kernel matters–Improve semantic segmentation by global convolutional network
ArXiv Preprint ArXiv:1703.02719
(2017)
Google Scholar ↗

Pfister et al., 2015 T. Pfister, J. Charles, A. Zisserman
Flowing ConvNets for human pose estimation in videos
ArXiv:1506.02897 [Cs]. Retrieved from http://arxiv.org/abs/1506.02897
(2015)
Google Scholar ↗

Priya et al., 2015 G. Priya, S.N. Paul, Y.J. Singh
Human walking motion detection and classification of actions from Video Sequences, 3 (1) (2015), p. 6
Google Scholar ↗

Quercia et al., 2014 D. Quercia, N.K. O'Hare, H. Cramer
Aesthetic capital: What makes london look beautiful, quiet, and happy?
Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW' 14 (2014), pp. 945-955, 10.1145/2531602.2531613 ↗
View in Scopus ↗ Google Scholar ↗

Radford et al., 2015 A. Radford, L. Metz, S. Chintala

Unsupervised representation learning with deep convolutional generative adversarial networks
ArXiv:1511.06434 [Cs]. Retrieved from
(2015)
http://arxiv.org/abs/1511.06434 ↗
Google Scholar ↗

Redmon et al., 2016 J. Redmon, S. Divvala, R. Girshick, A. Farhadi
You only look once: Unified, real-time object detection
ArXiv:1506.02640 [Cs]. Retrieved from
(2016)
http://arxiv.org/abs/1506.02640 ↗
Google Scholar ↗

Redmon and Farhadi, 2017 J. Redmon, A. Farhadi
YOLO9000: Better, faster, stronger
2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017), pp. 6517-6525, 10.1109/CVPR.2017.690 ↗
View in Scopus ↗ Google Scholar ↗

Redmon and Farhadi, 2018 J. Redmon, A. Farhadi
YOLOv3: An Incremental Improvement (2018), p. 6
Google Scholar ↗

Reed, Akata, Yan, et al., 2016 S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, H. Lee
Generative adversarial text to image synthesis
ArXiv:1605.05396 [Cs]. Retrieved from
(2016)
http://arxiv.org/abs/1605.05396 ↗
Google Scholar ↗

Reed, Akata, Mohan, et al., 2016 S.E. Reed, Z. Akata, S. Mohan, S. Tenka, B. Schiele, H. Lee
Learning what and where to draw
(2016), p. 9
Google Scholar ↗

Reichstein et al., 2019 M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, Prabhat
Deep learning and process understanding for data-driven Earth system science
Nature, 566 (7743) (2019), pp. 195-204, 10.1038/s41586-019-0912-1 ↗
View in Scopus ↗ Google Scholar ↗

Ren et al., 2016 S. Ren, K. He, R. Girshick, J. Sun
Faster R-CNN: Towards real-time object detection with region proposal networks
arXiv:1506.01497v3
(2016)
Google Scholar ↗

Robie et al., 2017 A.A. Robie, K.M. Seagraves, S.E.R. Egnor, K. Branson
Machine vision methods for analyzing social interactions
The Journal of Experimental Biology, 220 (1) (2017), pp. 25-34, 10.1242/jeb.142281 ↗
View in Scopus ↗ Google Scholar ↗

Ronneberger et al., 2015 O. Ronneberger, P. Fischer, T. Brox
U-net: Convolutional networks for biomedical image segmentation
ArXiv:1505.04597 [Cs]. Retrieved from
(2015)
http://arxiv.org/abs/1505.04597 ↗
Google Scholar ↗

Russakovsky et al., 2015 O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, ..., L. Fei-Fei
ImageNet large scale visual recognition challenge
International Journal of Computer Vision, 115 (3) (2015), pp. 211-252, 10.1007/s11263-015-0816-y ↗
Google Scholar ↗

Saha et al., 2017 S. Saha, G. Singh, F. Cuzzolin

AMTnet: Action-micro-Tube regression by end-to-end trainable deep architecture

ArXiv:1704.04952 [Cs]. Retrieved from

(2017)

http://arxiv.org/abs/1704.04952 ↗

Google Scholar ↗

Saha et al., 2016 S. Saha, G. Singh, M. Sapienza, P.H.S. Torr, F. Cuzzolin

Deep learning for detecting multiple space-time action tubes in videos

ArXiv:1608.01529 [Cs]. Retrieved from

(2016)

http://arxiv.org/abs/1608.01529 ↗

Google Scholar ↗

Salesses et al., 2013 P. Salesses, K. Schechtner, C.A. Hidalgo

The collaborative image of the city: Mapping the inequality of urban perception

PloS One, 8 (7) (2013), Article e68400, 10.1371/journal.pone.0068400 ↗

View in Scopus ↗ Google Scholar ↗

Sayed et al., 2013 T. Sayed, M.H. Zaki, J. Autey

Automated safety diagnosis of vehicle–bicycle interactions using computer vision analysis

Safety Science, 59 (2013), pp. 163-172, 10.1016/j.ssci.2013.05.009 ↗

View PDF View article View in Scopus ↗ Google Scholar ↗

Seresinhe et al., 2017 C.I. Seresinhe, T. Preis, H.S. Moat

Using deep learning to quantify the beauty of outdoor places

Royal Society Open Science, 4 (7) (2017), Article 170170, 10.1098/rsos.170170 ↗

View in Scopus ↗ Google Scholar ↗

Sharma et al., 2017 A. Sharma, X. Liu, X. Yang, D. Shi

A patch-based convolutional neural network for remote sensing image classification

Neural Networks, 95 (2017), pp. 19-28, 10.1016/j.neunet.2017.07.017 ↗

View PDF View article View in Scopus ↗ Google Scholar ↗

Shou et al., 2017 Z. Shou, J. Chan, A. Zareian, K. Miyazawa, S.-F. Chang

CDC: Convolutional-de-Convolutional networks for precise temporal action localization in untrimmed videos

ArXiv:1703.01515 [Cs]. Retrieved from

(2017)

http://arxiv.org/abs/1703.01515 ↗

Google Scholar ↗

Simonyan and Zisserman, 2014 K. Simonyan, A. Zisserman

Very deep convolutional networks for large-scale image recognition

ArXiv Preprint ArXiv:1409.1556

(2014)

Google Scholar ↗

Singh et al., 2016 G. Singh, S. Saha, M. Sapienza, P. Torr, F. Cuzzolin

Online real-time multiple spatiotemporal action localisation and prediction

ArXiv:1611.08563 [Cs]. Retrieved from

(2016)

http://arxiv.org/abs/1611.08563 ↗

Google Scholar ↗

Sirirattanapol et al., 2019 C. Sirirattanapol, M. Nagai, A. Witayangkurn, S. Pravinvongvuth, M. Ekpanyapong

Bangkok CCTV image through a road environment extraction system using multi-label convolutional neural network classification

ISPRS International Journal of Geo-information, 8 (3) (2019), p. 128, 10.3390/ijgi8030128 ↗

View in Scopus ↗ Google Scholar ↗

Soomro and Shah, 2017 K. Soomro, M. Shah

Unsupervised action Discovery and localization in videos.

2017 IEEE International Conference on Computer Vision (ICCV) (2017), pp. 696-705, 10.1109/ICCV.2017.82 ↗

View in Scopus ↗ Google Scholar ↗

Srivastava et al., 2019 S. Srivastava, J.E. Vargas-Muñoz, D. Tuia

Understanding urban landuse from the above and ground perspectives: A deep learning, multimodal solution

Remote Sensing of Environment, 228 (2019), pp. 129-143, 10.1016/j.rse.2019.04.014 ↗

View PDF View article View in Scopus ↗ Google Scholar ↗

Sun et al., 2017 Y. Sun, Y. Liu, G. Wang, H. Zhang

Deep learning for plant identification in natural environment

Computational Intelligence and Neuroscience, 2017 (2017), pp. 1-6, 10.1155/2017/7361042 ↗

View PDF View article Google Scholar ↗

Szegedy et al., 2015 C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed

Going deeper with convolutions

Retrieved from https://www.cs.unc.edu/~wliu/papers/GoogLeNet.pdf

(2015)

Google Scholar ↗

Tian et al., 2017 K. Tian, S. Zhou, J. Guan

DeepCluster: A General clustering framework based on deep learning

M. Ceci, J. Hollmén, L. Todorovski, C. Vens, S. Džeroski (Eds.), Machine learning and knowledge Discovery in databases, Vol. 10535 (2017), pp. 809-825, 10.1007/978-3-319-71246-8_49 ↗

View in Scopus ↗ Google Scholar ↗

van Hasselt et al., 2015v H. van Hasselt, A. Guez, D. Silver

Deep reinforcement learning with double Q-Learning

(2015), p. 7

Google Scholar ↗

van Veenstra and Kotterink, 2017 A.F. van Veenstra, B. Kotterink

Data-driven policy making: The policy lab approach

P. Parycek, Y. Charalabidis, A.V. Chugunov, P. Panagiotopoulos, T.A. Pardo, Ø. Sæbø, E. Tambouris (Eds.), Electronic participation, Springer International Publishing. (2017), pp. 100-111

CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Vanhoey et al., 2017 K. Vanhoey, D. Dai, L. Van Gool, C.E.P. de Oliveira, H. Riemenschneider, A. Bódis-Szomorú, ..., T. Kroeger

VarCity - The video: The struggles and triumphs of leveraging fundamental research results in a graphics video production

ACM SIGGRAPH 2017 Talks on - SIGGRAPH' 17 (2017), pp. 1-2, 10.1145/3084363.3085085 ↗

Google Scholar ↗

Viola and Jones, 2001 P. Viola, M. Jones

Rapid object detection using a boosted cascade of simple features. Computer vision and pattern recognition, 2001. CVPR 2001

Proceedings of the 2001 IEEE Computer Society Conference On, 1, I–I. IEEE (2001)

Google Scholar ↗

Voigtlaender et al., 2019 P. Voigtlaender, M. Krause, B.B.G. Sekar, A. Geiger, B. Leibe

MOTS: Multi-object tracking and segmentation

(2019), p. 10

Google Scholar ↗

Wang, Zhou, et al., 2018 B. Wang, W. Zhao, P. Gao, Y. Zhang, Z. Wang

Crack damage detection method via multiple visual features and efficient multi-task learning model

Sensors, 18 (6) (2018), p. 1796, 10.3390/s18061796 ↗

View in Scopus ↗ Google Scholar ↗

Wang, Xu, et al., 2018 L. Wang, X. Xu, H. Dong, R. Gui, F. Pu

Multi-pixel simultaneous classification of PolSAR image using convolutional neural networks

Sensors, 18 (3) (2018), p. 769, 10.3390/s18030769 ↗

View in Scopus ↗ Google Scholar ↗

Wang, Qiao et al., 2015 L. Wang, Y. Qiao, X. Tang
Action recognition with trajectory-pooled deep-convolutional descriptors
2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015), pp. 4305-4314, 10.1109/CVPR.2015.7299059 ↗
View in Scopus ↗ Google Scholar ↗

Wang, Quan, et al., 2018 S. Wang, D. Quan, X. Liang, M. Ning, Y. Guo, L. Jiao
A deep learning framework for remote sensing image registration
ISPRS Journal of Photogrammetry and Remote Sensing (2018), 10.1016/j.isprsjprs.2017.12.012 ↗
Google Scholar ↗

Wang, Yang, et al., 2018 W. Wang, S. Yang, Z. He, M. Wang, J. Zhang, W. Zhang
Urban perception of commercial activeness from satellite images and streetscapes.
Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW' 18 (2018), pp. 647-654, 10.1145/3184558.3186581 ↗
View in Scopus ↗ Google Scholar ↗

Wang, Schaul, et al., 2015 Z. Wang, T. Schaul, M. Hessel, H. van Hasselt, M. Lanctot, N. de Freitas
Dueling network architectures for deep reinforcement learning
ArXiv:1511.06581 [Cs]. Retrieved from
(2015)
http://arxiv.org/abs/1511.06581 ↗
Google Scholar ↗

Weinzaepfel et al., 2015 P. Weinzaepfel, Z. Harchaoui, C. Schmid
Learning to track for spatio-temporal action localization
2015 IEEE International Conference on Computer Vision (ICCV) (2015), pp. 3164-3172, 10.1109/ICCV.2015.362 ↗
View in Scopus ↗ Google Scholar ↗

Weinzaepfel et al., 2016 P. Weinzaepfel, X. Martin, C. Schmid
Human action localization with sparse spatial supervision
ArXiv:1605.05197 [Cs]. Retrieved from
(2016)
http://arxiv.org/abs/1605.05197 ↗
Google Scholar ↗

Williams et al., 2017 D. Williams, A. Britten, S. McCallum, H. Jones, M. Aitkenhead, A. Karley, ..., J. Graham
A method for automatic segmentation and splitting of hyperspectral images of raspberry plants collected in field conditions
Plant Methods, 13 (1) (2017), 10.1186/s13007-017-0226-y ↗
Google Scholar ↗

Wu et al., 2016 G. Wu, W. Lu, G. Gao, C. Zhao, J. Liu
Regional deep learning model for visual tracking
Neurocomputing, 175 (2016), pp. 310-323, 10.1016/j.neucom.2015.10.064 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Wurm et al., 2019 M. Wurm, T. Stark, X.X. Zhu, M. Weigand, H. Taubenböck
Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks
ISPRS Journal of Photogrammetry and Remote Sensing, 150 (2019), pp. 59-69, 10.1016/j.isprsjprs.2019.02.006 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Xie et al., 2016 J. Xie, R. Girshick, A. Farhadi
Unsupervised deep embedding for clustering analysis
(2016), p. 10
CrossRef ↗ Google Scholar ↗

Xu et al., 2017 H. Xu, A. Das, K. Saenko
R-C3D: Region convolutional 3D network for temporal activity detection
ArXiv:1703.07814 [Cs]. Retrieved from
(2017)
http://arxiv.org/abs/1703.07814 ↗

Google Scholar ↗

Yang et al., 2019 D. Yang, X. Liu, H. He, Y. Li
Air-to-ground multimodal object detection algorithm based on feature association learning
International Journal of Advanced Robotic Systems, 16 (3) (2019), Article 172988141984299, 10.1177/1729881419842995 ↗
Google Scholar ↗

Yang et al., 2018 M. Yang, K. Yu, C. Zhang, Z. Li, K. Yang
DenseASPP for semantic segmentation in street scenes
(2018), p. 9
Google Scholar ↗

Yang and Pun-Cheng, 2018 Z. Yang, L.S.C. Pun-Cheng
Vehicle detection in intelligent transportation systems and its applications under varying environments: A review
Image and Vision Computing, 69 (2018), pp. 143-154, 10.1016/j.imavis.2017.09.008 ↗
View PDF View article Google Scholar ↗

Yu and Koltun, 2015 F. Yu, V. Koltun
Multi-scale context aggregation by dilated convolutions
ArXiv Preprint ArXiv:1511.07122
(2015)
Google Scholar ↗

Yu et al., 2017 H. Yu, Z. Wu, S. Wang, Y. Wang, X. Ma
Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks
Sensors, 17 (12) (2017), p. 1501, 10.3390/s17071501 ↗
View in Scopus ↗ Google Scholar ↗

Zaki and Sayed, 2013 M.H. Zaki, T. Sayed
A framework for automated road-users classification using movement trajectories
Transportation Research Part C, Emerging Technologies, 33 (2013), pp. 50-73, 10.1016/j.trc.2013.04.007 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Zaki et al., 2013 M.H. Zaki, T. Sayed, A. Tageldin, M. Hussein
Application of computer vision to diagnosis of pedestrian safety issues
Transportation Research Record: Journal of the Transportation Research Board, 2393 (1) (2013), pp. 75-84, 10.3141/2393-09 ↗
View in Scopus ↗ Google Scholar ↗

Zhang, Wang, et al., 2016 B. Zhang, L. Wang, Z. Wang, Y. Qiao, H. Wang
Real-time action recognition with enhanced motion vector CNNs
ArXiv:1604.07669 [Cs]. Retrieved from
(2016)
http://arxiv.org/abs/1604.07669 ↗
Google Scholar ↗

Zhang et al., 2019 F. Zhang, L. Wu, D. Zhu, Y. Liu
Social sensing from street-level imagery: A case study in learning spatio-temporal urban mobility patterns
ISPRS Journal of Photogrammetry and Remote Sensing, 153 (2019), pp. 48-58, 10.1016/j.isprsjprs.2019.04.017 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Zhang, Zhou, et al., 2018 F. Zhang, B. Zhou, L. Liu, Y. Liu, H.H. Fung, H. Lin, C. Ratti
Measuring human perceptions of a large-scale urban region using machine learning
Landscape and Urban Planning, 180 (2018), pp. 148-160, 10.1016/j.landurbplan.2018.08.020 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Zhang, Xu, et al., 2016 H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, D. Metaxas
StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks
ArXiv:1612.03242 [Cs, Stat]. Retrieved from
(2016)
http://arxiv.org/abs/1612.03242 ↗
Google Scholar ↗

Zhang et al., 2017 S. Zhang, Z. Wei, J. Nie, L. Huang, S. Wang, Z. Li

A review on human activity recognition using vision-based method
Journal of Healthcare Engineering, 2017 (2017), pp. 1-31, 10.1155/2017/3090343 ↗
View in Scopus ↗ Google Scholar ↗

Zhang, Xia, et al., 2018 X. Zhang, G.-S. Xia, Q. Lu, W. Shen, L. Zhang
Visual object tracking by correlation filters and online learning
ISPRS Journal of Photogrammetry and Remote Sensing, 140 (2018), pp. 77-89, 10.1016/j.isprsjprs.2017.07.009 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Zhao, Shi, et al., 2017 H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia
Pyramid scene parsing network
IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2017), pp. 2881-2890
Google Scholar ↗

Zhao et al., 2018 J. Zhao, X. Liu, Y. Kuang, Y.V. Chen, B. Yang
Deep CNN-based methods to evaluate neighborhood-scale Urban valuation through Street scenes perception
2018 IEEE Third International Conference on Data Science in Cyberspace (DSC) (2018), pp. 20-27, 10.1109/DSC.2018.00012 ↗
View PDF View article Google Scholar ↗

Zhao, Xiong, et al., 2017 Y. Zhao, Y. Xiong, L. Wang, Z. Wu, X. Tang, D. Lin
Temporal action detection with structured segment networks (2017)
ArXiv:1704.06228 [Cs]. Retrieved from
http://arxiv.org/abs/1704.06228 ↗
Google Scholar ↗

Zhou, Zhao, et al., 2017 B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, A. Torralba
Scene parsing through ADE20K dataset
2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017), pp. 5122-5130, 10.1109/CVPR.2017.544 ↗
View in Scopus ↗ Google Scholar ↗

Zhu, Vial, et al., 2017 H. Zhu, R. Vial, S. Lu
TORNADO: A spatio-temporal convolutional regression network for video action proposal
2017 IEEE International Conference on Computer Vision (ICCV) (2017), pp. 5814-5822, 10.1109/ICCV.2017.619 ↗
View in Scopus ↗ Google Scholar ↗

Zhu, Lan, et al., 2017 Y. Zhu, Z. Lan, S. Newsam, A.G. Hauptmann
Hidden Two-stream convolutional networks for action recognition (2017)
ArXiv:1704.00389 [Cs]. Retrieved from
http://arxiv.org/abs/1704.00389 ↗
Google Scholar ↗

Zou et al., 2019 Z. Zou, Z. Shi, Y. Guo, J. Ye
Object detection in 20 years: A survey
ArXiv:1905.05055v2
(2019), p. 40
Google Scholar ↗

Cited by (106)

Cooling effects in urban communities: Parsing green spaces and building shadows
2024, Urban Forestry and Urban Greening
Show abstract

An integrated deep learning approach for assessing the visual qualities of built environments utilizing street view images
2024, Engineering Applications of Artificial Intelligence
Show abstract

Towards safer streets: A framework for unveiling pedestrians’ perceived road safety using street view imagery
2024, Accident Analysis and Prevention

Show abstract

Deep hybrid model with satellite imagery: How to combine demand modeling and computer vision for travel behavior analysis?

2024, Transportation Research Part B: Methodological

Show abstract

Mapping property redevelopment via GeoAI: Integrating computer vision and socioenvironmental patterns and processes

2024, Cities

Show abstract

Deep learning video analytics for the assessment of street experiments: The case of Bologna

2023, Journal of Urban Mobility

Show abstract

View all citing articles on Scopus ↗

View Abstract

ELSEVIER

RELX™