Snakes in Trees: An Explainable Artificial Intelligence Approach for Automatic Object Detection and Recognition

Joanna Isabelle Olszewska
School of Computing and Engineering, University of the West of Scotland, U.K.

Keywords: Explainable Artificial Intelligence, Explainable by Design, Computer Vision, Machine Vision, Smart Cities, Industry 4.0, Intelligent Systems, Decision Tree, Snake, Active Contours, Recursive Algorithm, Unsupervised Labeling, Semantic Tag, Automatic Image Annotation.

Abstract: Nowadays, the development of smart cities boosts the development of innovative IT technologies based on Artificial Intelligence (AI), such as intelligent agents (IA), which themselves use new algorithms, complex software, and advanced systems. However, due to their expanding number and range of applications as well as their growing autonomy, there is an increased expectation for these intelligent technologies to involve explainable algorithms, dependable software, trustworthy systems, transparent agents, etc. Hence, in this paper, we present a new explainable algorithm which uses snakes within trees to automatically detect and recognize objects. The proposed method involves the recursive computation of snakes (aka parametric active contours), leading to multi-layered snakes where the first layer corresponds to the main object of interest, while the next-layer snakes delineate the different sub-parts of this foreground. Visual features are extracted from the regions segmented by these snakes and are mapped into semantic concepts. Based on these attributes, decision trees are induced, resulting in effective semantic labeling of the objects and the automatic annotation of the scene. Our computer-vision approach shows excellent computational performance on real-world standard database, in context of smart cities.

1 INTRODUCTION

According to the United Nations (UN), more than half of the World population currently lives in urban areas, and this trend is only rising (Bhattacharya et al., 2020). Therefore, there is a need to rethink cities in efficient and modern ways, leading to *smart cities*, which are defined as urban areas that create sustainable economic development and high quality of life by excelling in six key areas, namely, economy, mobility, environment, people, living, and government (Montemayor et al., 2015).

In smart cities, traditional infrastructures as well as new services are merged, coordinated, and integrated using innovative digital technologies (Batty et al., 2012). Hence, IT technologies such Artificial Intelligence (AI)-based systems are a cornerstone aspect of smart cities, with AI acting as a catalyst for areas such as smart urban modeling, intelligent infrastructures, smart transportation, smart governance, sustainability, smart education, and smart health solutions, to name a few (Bhattacharya et al., 2020).

In particular, computer-vision-based intelligent systems contribute to smart city's applications such as smart surveillance (Ryabchikov et al., 2020), smart security (Khan et al., 2019), smart traffic management (Gupta and Sundar, 2020), and smart mobility (Fortes et al., 2021). Intelligent vision systems can thus improve people's quality of life by helping drivers with finding vacant parking space (Bravo et al., 2013) or aiding visually impaired to safely travel around the smart city (Nasralla et al., 2019).

Intelligent vision systems are also embedded in a variety of intelligent agents (Olszewska, 2020) such as inspection robots (Sui, 2021), cleaning robots (Narang et al., 2021), etc. They can also be integrated into mobile ground robots for assistive operations in public spaces (Grzeskowiak et al., 2021) or into unmanned aerial vehicles (UAV) for data collection through the smart city (Shirazi et al., 2020).

All these computer-vision-based applications involve automatic object detection (Chen et al., 2021). Indeed, automatic object detection is used for object recognition (Wang et al., 2021), scene recognition (Zeng et al., 2020), activity recognition (Mliki et al.,

Olszewska, J.
Snakes in Trees: An Explainable Artificial Intelligence Approach for Automatic Object Detection and Recognition.
DOI: 10.5220/0010993000003116
In *Proceedings of the 14th International Conference on Agents and Artificial Intelligence (ICAART 2022) - Volume 3*, pages 996-1002
ISBN: 978-989-758-547-0; ISSN: 2184-433X

2020), object tracking (Li et al., 2021), or automated image annotation (Zhang et al., 2012), which in turn can be applied to pedestrian detection and surveillance (An et al., 2021), car detection and annotation (Li et al., 2020), license plate recognition (Huang et al., 2021), robot navigation (Lin et al., 2021), and many more applications (Calzado et al., 2018), (Olszewska, 2018).

Therefore, the development of transparent and explainable artificial intelligence algorithms for intelligent vision systems (Olszewska, 2021) deployed in smart cities is of prime importance, due to the growing concern of citizens about, on one hand, their data privacy and security (Chourabi et al., 2012), and, on the other hand, the behaviour of the surrounding intelligent agents (Zhang et al., 2017).

Over the last decade, computer vision systems have been more and more relying on machine learning, and especially on deep learning (Camacho et al., 2021), which is a very popular and efficient approach. However, the use of deep learning involves most of the time very large and unbiased training datasets (Kishida et al., 2021) with high-resolution images (Wang et al., 2021), time and energy-consuming processes (Liu et al., 2021), resource-intensive computational power (Prakash et al., 2020), with associated, extensive costs (Chourabi et al., 2012) as well as sophisticated equipment (Namiki et al., 2021); all these demands being not always available in real-world conditions. Besides, deep learning is not currently considered as an explainable machine-learning approach (Gunning et al., 2019).

Since explainable artificial intelligence (XAI) is very important for users (Ferreira, J. J. and Monteiro, M. S., 2020), (Wilding et al., 2020) and regulators (Winfield et al., 2021), (Prestes et al., 2021), explainability becomes a 'sought-after' non-functional requirement (Kohl et al., 2019) of AI-based systems such as intelligent vision systems (Olszewska, 2019b).

Explainability can thus be assessed by internal algorithmic auditing (Raji et al., 2020), software testing (Black et al., 2022), and/or by verification and validation of the intelligent system (Corso et al., 2021).

Explainability can be also addressed at an earlier stage of the intelligent-vision system development (Olszewska, 2019a), i.e. during the system design (Bryson and Winfield, 2017) and its algorithm design (Mendling et al., 2021), leading to *XAI by Design* (Kearns and Roth, 2020).

Thus, in this work, we propose a novel, explainable-by-design AI-based algorithm for intelligent vision systems. Our algorithm consists mainly on the recursive computation of both the computer-vision method called *snakes* and the machine-learning-based *decision trees*.

Indeed, on one hand, object-of-interest's shape (Samani et al., 2021) and closed contours (Funke et al., 2021) are very important visual feature for object detection and recognition (Lv et al., 2021). Therefore, computer-vision techniques such as active contours (Yezzi et al., 2019) are an efficient and explainable method to locate, segment, and track an object. In particular, we have selected the active-contour method known as 'snake' (Muralidhar et al., 2010), which automatically computes parametric active contours to delineate the visual-object shape, since snakes ally explainability with excellent detection performance (Olszewska, 2017).

On the other hand, we have adopted decision trees (DT), which are considered as the most explainable approach to machine learning (Gunning et al., 2019). Furthermore, decision trees are efficient for visual object detection and recognition applications (Nowozin et al., 2011), (Olszewska, 2015b).

Thence, our explainable-by-design algorithm can perform object detection and recognition in both static or dynamic scenes, depending of the type of input data - still image (Li et al., 2020) or video frame (Wang et al., 2019), respectively.

Our algorithm can also process different levels of granularity, which is important for robust object detection and complex scene description (Wang et al., 2014).

Moreover, snakes allow object recognition with an open-set domain (Kishida et al., 2021), without requiring any cumbersome training.

Besides, snakes can be coupled with ontologies such as the Spatio-Temporal Visual Ontology (STVO) (Olszewska and McCluskey, 2011), and therefore, they can directly bridge vision systems and knowledge-based systems. Furthermore, through STVO, they can be connected to other ontologies which are suitable for cutting-edge vision-embedded technologies such as autonomous systems (Olivares-Alarcos et al., 2019), cloud-robotic systems (Pignaton de Freitas et al., 2020), smart manufacturing (Hildebrandt et al., 2020), or smart cities (Burns et al., 2018).

Hence, all this algorithmic design leads to a transparent and efficient visual object detection and automated semantic scene annotation, while provides an explainable and energy-efficient solution for intelligent vision agents to be deployed in smart cities.

Thus, the contribution of this paper is the new, explainable algorithm that allies recursive, multi-layered snakes with recursive, decision trees for machine-vision object detection and recognition.

The paper is structured as follows. In Section 2, we present our XAI approach for automatic, visual object detection and recognition as well as unsupervised, semantic labeling of an image by means of multi-layer, recursive snakes within induced decision trees. The resulting annotation system has been successfully tested on a challenging database containing real-world images as reported and discussed in Section 3. Conclusions are drawn up in Section 4.

2 OUR PROPOSED APPROACH

In this section, we present our 'Snakes-in-Trees' (ST) approach, which allows both the automatic visual object detection and recognition as well as its automatic semantic labeling, as exemplified in Fig. 1.

For this purpose, our ST algorithm (Algorithm 1) computes recursively snakes (S) in an input image (I), in order to detect an object (O) in a robust way against occlusions and to map the object's semantic label (L) granularity for a precise object recognition and image annotation.

In this work, each snake S is implemented by a multi-feature active contour (Olszewska, 2015a) which is defined as a parametric curve $\mathcal{C}(s) : [0,1] \to \mathbb{R}^2$ modeled by a B-spline formalism and guided by both the internal forces (α: elasticity, β: rigidity) resulting from the curve's mechanical properties and the external force Ξ resulting from multiple features of the input image, as per following dynamic equation:

$$\mathcal{C}_t(s,t) = \alpha\, \mathcal{C}_{ss}(s,t) - \beta\, \mathcal{C}_{ssss}(s,t) + \Xi. \quad (1)$$

The recursive computation of snakes $S_{l+i,k}$ within the input image I is performed by applying $ith-$times the Eq. (1) to I and leads to the multi-layered $(l+i)$ partition of I in terms of object-of-interest's background (at the layer $l+i = l$, with $i = 0$), its foreground (at the layer $l+i = l+1$, with $i = 1$) as well as the foregrounds of semantically meaningful subobjects (at the subsequent layers $l+i$, with $i \geq 2$) of the object of interest delineated by k snakes at the corresponding layers $l+i$.

Hence, this recursive process enables the automatic detection of coherent and consistent visual objects, which are described by geometric and metric properties (Olszewska, 2013). These features serve to the characterization of the regions extracted by the snakes and contribute to define the objects or their subparts in terms of both numeric and semantic concepts.

The latter ones are recursively mapped into natural-language keywords through the pre-order traversal of rooted trees that are recursively computed by the ST algorithm, allowing the efficient object labeling as well as accurate object recognition (see Algorithm 1).

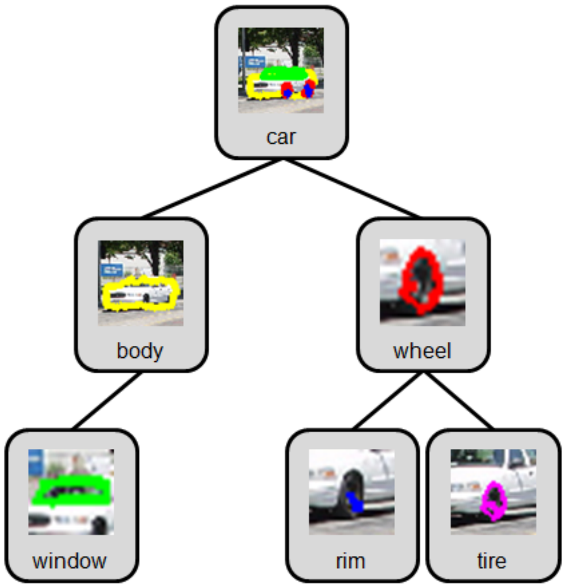

Figure 1: Overview of our 'Snakes-in-Trees' process for the 'car' tag object.

It is worth noting that the decision trees, which are recursively built by the ST algorithm, ensure a granular and semantic mapping of the visual objects that are detected in each layer by the snakes, which are themselves recursively computed by the ST algorithm, for an accurate object detection, recognition, and annotation.

Hence, decision trees are induced in order to define semantic keywords at each level corresponding to visual feature defined by each layer of the snakes. Then, a voting mechanism allows for higher semantic level decisions in order to recognize the object.

3 RESULTS AND EVALUATION

To validate our transparent algorithmic method for the automatic visual object detection and recognition as well as automatic image annotation in context of smart cities, we used the publicly available MIT CBCL street scenes database, which contains 35,417 jpg images with a resolution of 1280x960.

We carried out experiments consisting in running our recursive algorithm implemented in MatLab on a commercial device with a processor Intel(R) Core(TM)2 Duo CPU T9300 2.50 GHz, 2 Gb RAM and the MatLab (Mathworks, Inc.) and applied to the CBCL database.

(a) (b) (c) (d)

Figure 2: Examples of results of our recursive ‘Snakes-in-Trees’ method, when applied to the CBCL dataset. Best viewed in color.

Table 1: Average accuracy of object-of-interest recognition, using approaches of △(Ren et al., 2015), ◊(Lu et al., 2019), ○(Liu et al., 2008), □(Kim et al., 2004), and our ‘Snakes-in-Trees’ (ST) method.

method	△	◊	○	□	ST
average accuracy	65.4%	69.5%	73.1%	84.2%	95.6%

Algorithm 1: ST: Snakes in Trees.

Given O, the object of interest to tag;
Given T_O, the Tree related to the object O;
Given $S_{l+i,k}$, the kth Snake at the $l+ith$ layer;
Given $L_{O,n}$, the label of the root n of the tree T_O;

Considering
$T_{O,l+i}$, the $l+ith$ level of the tree T_O;
$i=0$;
$V=\varnothing$;
$r_{S_{l+i,k}}=0$;

ST($S_{l+i,k}, T_O$)
  $L_{O,n}$ =label(head($T_{O,l+i}$))
  $S_{l+i,k+m} \leftarrow S_{l+i,k}$
  $S_{l+i,k} = S_{l+i,k+m}$
  if $\exists S_{l+i,k} | (S_{l+i,k} \vDash L_{O,n})$ **then**
    $r_{S_{l+i,k}} = 1$
    $V = V \cup \{r_{S_{l+i,k}}\}$
  end if

  if $T_{O,l+i,left} \neq \varnothing$ **then**
    $i = i+1$
    ST($S_{l+i,k}$,$T_{O,l+i,left}$)
  end if

  if $T_{O,l+i,right} \neq \varnothing$ **then**
    $i = i+1$
    ST($S_{l+i,k}$,$T_{O,l+i,right}$)
  end if

  if MajorityVote(V) == 1 **then**
    return label(head(T_O))
  end if

end

A sample of obtained results, which have been computed in few ms, can be found in Figs. 2 (a)-(d). We can observe that the algorithm has well detected and recognized the objects of interest as well as their components, delineating and labeling them appropriately.

To quantitatively assess the accuracy of our algorithm, we adopt the standard criterion (Olszewska, 2019b), as follows:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad (2)$$

with TP, true positive, TN, true negative, FP, false positive, and FN, false negative.

Table 1 reports the mean average accuracy, which is computed as per Eq. (2) in the case of our algorithm (i.e. 95.6%), and compares it to the performance of other state-of-the-art machine-learning methods for object detection and recognition. On one hand, we can note that current methods (Wang et al., 2021) based on deep learning and convolutional neural networks (Fathima and Merriliance, 2020), such as Faster R-CNN (Ren et al., 2015) and Grid R-CNN (Lu et al., 2019), achieve only an accuracy of 65.4% and 69.5%, respectively. As reported by (Khan et al., 2020), CBCL dataset images are challenging because of their low resolution for training as well as testing neural networks (Jimenez-Bravo et al., 2020). On the other hand, common decision-tree-based methods, such as (Liu et al., 2008) and (Kim et al., 2004), have an accuracy of 73.1% and 84.2%, respectively, and they involve constraining assumptions such as recognition of very distant classes only or detection of only centered objects of interest, respectively. Therefore, our ST algorithm features not only explainability, but it is also low-spec, cost-effective, time-efficient, and robust, while being more accurate in comparison with state-of-the-art approaches.

4 CONCLUSIONS

In this paper, we propose an explainable-by-design algorithm built on snakes within trees for automatic object detection and recognition. Indeed, we have developed an efficient XAI algorithm embedding recursive snakes within the recursive pre-order traversal of rooted trees, where each decision tree's semantic value has been mapped with the visual information provided by a layer of the computed snake. Hence, based on both semantic and visual properties of the image content, our 'Snakes-in-Trees' (ST) method provides accurate and robust object detection and recognition as well as image annotation in real-world conditions, even in case of pose variations or occlusions of the objects of interest. The ST algorithm is thus well suited for smart cities' intelligent-vision-based applications.

REFERENCES

An, H., Hu, H.-M., Guo, Y., Zhou, Q., and Li, B. (2021). Hierarchical reasoning network for pedestrian attribute recognition. *IEEE Transactions on Multimedia*, 23:268–280.

Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., and Portugali, Y. (2012). Smart cities of the future. *The European Physical Journal Special Topics*, 214(1):481–518.

Bhattacharya, S., Somayaji, S. R. K., Gadekallu, T. R., Alazab, M., and Maddikunta, P. K. R. (2020). A review on deep learning for future smart cities. *Internet Technology Letters*, pages 1–6.

Black, R., Davenport, J. H., Olszewska, J. I., Roessler, J., Smith, A. L., and Wright, J. (2022). *Artificial Intelligence and Software Testing: A Practical Guide to Quality*. BCS Press.

Bravo, C., Sanchez, N., Garcia, N., and Menendez, J. M. (2013). Outdoor vacant parking space detector for improving mobility in smart cities. In *Proceedings of the Portuguese Conference on Artificial Intelligence*, pages 30–41.

Bryson, J. and Winfield, A. (2017). Standardizing ethical design for artificial intelligence and autonomous systems. *IEEE Computer*, 50(5):116–119.

Burns, M., Griffor, E., Balduccini, M., Vishik, C., Huth, M., and Wollman, D. (2018). Reasoning about smart city. In *Proceedings of the IEEE International Conference on Smart Computing*, pages 381–386.

Calzado, J., Lindsay, A., Chen, C., Samuels, G., and Olszewska, J. I. (2018). SAMI: Interactive, multi-sense robot architecture. In *Proceedings of the IEEE International Conference on Intelligent Engineering Systems*, pages 317–322.

Camacho, A., Varley, J., Zeng, A., Jain, D., Iscen, A., and Kalashnikov, D. (2021). Reward machines for vision-based robotic manipulation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 14284–14290.

Chen, Z.-M., Jin, X., Zhao, B.-R., Zhang, X., and Guo, Y. (2021). HCE: Hierarchical context embedding for region-based object detection. *IEEE Transactions on Image Processing*, 30:6917–6929.

Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J. R., Mellouli, S., Nahon, K., Pardo, T. A., and Scholl, H. J. (2012). Understanding smart cities: An integrative framework. In *Proceedings of the IEEE Hawaii International Conference on System Sciences*, pages 2289–2297.

Corso, A., Moss, R. J., Koren, M., Lee, R., and Kochenderfer, M. J. (2021). A survey of algorithms for black-box safety validation of cyber-physical systems. *Journal of Artificial Intelligence Research*, 72:377–428.

Fathima, S. A. and Merriliance, K. (2020). Analysis of vehicle detection using region-based convolutional neural networks (RCNN). *Journal of Xi'an University of Architecture and Technology*, 12(7):81–90.

Ferreira, J. J. and Monteiro, M. S. (2020). What are people doing about XAI user experience? A survey on AI explainability research and practice. In *Proceedings of HCI International Conference (HCI). LNCS 12201, Part II*, pages 56–73.

Fortes, S., Kulesza, R., and Li, J. J. (2021). A case study of object recognition from drone videos. In *Proceedings of the IEEE International Conference on Information and Computer Technologies*, pages 84–87.

Funke, C. M., Borowski, J., Stosio, K., Brendel, W., Wallis, T. S. A., and Bethge, M. (2021). Five points to check when comparing visual perception in humans and machines. *Journal of Vision*, 16(3):1–23.

Grzeskowiak, F., Gonon, D., Dugas, D., Paez-Granados, D., Chung, J. J., Nieto, J., Siegwart, R., Billard, A., Babel, M., and Pettre, J. (2021). Crowd against the machine: A simulation-based benchmark tool to evaluate and compare robot capabilities to navigate a human crowd. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 3879–3885.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., and Yang, G.-Z. (2019). XAI-Explainable artificial intelligence. *Science Robotics*, 4(37):1–4.

Gupta, S. and Sundar, B. (2020). A computer vision based approach for automated traffic management as a smart city solution. In *Proceedings of the IEEE International Conference on Electronics, Computing and Communication Technologies*, pages 1–6.

Hildebrandt, C., Kocher, A., Kustner, C., Lopez-Enriquez, C.-M., Muller, A. W., Caesar, B., Gundlach, C. S., and Fay, A. (2020). Ontology building for cyber-physical systems: Application in the manufacturing domain. *IEEE Transactions on Automation Science and Engineering*, 17(3):1266–1282.

Huang, Q., Cai, Z., and Lan, T. (2021). A new approach for character recognition of multi-style vehicle license plates. *IEEE Transactions on Multimedia*, 23:3768–3777.

Jimenez-Bravo, D. M., Mutombo, P. M., Braem, B., and Marquez-Barja, J. M. (2020). Applying Faster R-CNN in extremely low-resolution thermal images for people detection. In *Proceedings of the IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications*, pages 1–4.

Kearns, M. and Roth, A. (2020). Ethical algorithm design. *ACM SIGecom Exchanges*, 18(1):31–36.

Khan, G., Samyan, S., Khan, M. U. G., Shahid, M., and Wahla, S. Q. (2020). A survey on analysis of human faces and facial expressions datasets. *International Journal of Machine Learning and Cybernetics*, 11(3):553–571.

Khan, M. M., Ilyas, M. U., Saleem, S., Alowibdi, J. S., and Alkatheiri, J. S. (2019). Emerging computer vision based machine learning issues for smart cities. In *The International Research and Innovation Forum*, pages 315–322.

Kim, S., Park, S., and Kim, M. (2004). Image classification into object/non-object classes. In *Proceedings of the International Conference on Image and Video Retrieval*, pages 393–400.

Kishida, I., Chen, H., Baba, M., Jin, J., Amma, A., and Nakayama, H. (2021). Object recognition with continual open set domain adaptation for home robot. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1516–1525.

Kohl, M. A., Baum, K., Langer, M., Oster, D., Speith, T., and Bohlender, D. (2019). Explainability as a non-functional requirement. In *Proceedings of the IEEE International Requirements Engineering Conference*, pages 363–368.

Li, X., Ma, H., and Luo, X. (2020). Weaklier supervised semantic segmentation with only one image level annotation per category. *IEEE Transactions on Image Processing*, 290:128–141.

Li, Y., Peng, R., Li, M., and Fang, C. (2021). Research on foreground Object recognition tracking and background restoration in AIoT era. In *Proceedings of the IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers*, pages 292–298.

Lin, S., Wang, J., Xu, M., Zhao, H., and Chen, Z. (2021). Topology aware object-level semantic mapping towards more robust loop closure. *IEEE Robotics and Automation Letters*, 6(4):7041–7048.

Liu, L., Tang, J., Liu, S., Yu, B., Xie, Y., and Gaudiot, J.-L. (2021). P-RT: A runtime framework to enable energy-efficient, real-time robotic vision applications on heterogeneous architectures. *IEEE Computer*, 54(4):14–25.

Liu, Y., Zhang, D., and Lu, G. (2008). Region-based image retrieval with high-level semantics using decision tree learning. *Pattern Recognition*, 41(8):2554–2570.

Lu, X., Li, B., Yue, Y., Li, Q., and Yan, J. (2019). Grid R-CNN. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7363–7372.

Lv, H., Zhang, F., and Wang, R. (2021). Robust active contour model using patch-based signed pressure force and optimized fractional-order edge. *IEEE Access*, 9:8771–8785.

Mendling, J., Depaire, B., and Leopold, H. (2021). Theory and practice of algorithm engineering. pages 1–18.

Mliki, H., Bouhlel, F., and M., H. (2020). Human activity recognition from UAV-captured video sequences. *Pattern Recognition*, 100:1–13.

Montemayor, A. S., Pantrigo, J. J., and Salgado, L. (2015). Special issue on real-time computer vision in smart citiess. *Journal of Real-Time Image Processing*, 10:723–724.

Muralidhar, G. S., Bovik, A. C., Giese, J. D., Sampat, M. P., Whitman, G. J., Haygood, T. M., Stephens, T. W., and Markey, M. K. (2010). Snakules: A model-based active contour algorithm for the annotation of spicules on mammography. *IEEE Transactions on Medical Imaging*, 29(10):1768–1780.

Namiki, S., Yokoyama, K., Yachida, S., Shibata, T., Miyano, H., and Ishikawa, M. (2021). Online object recognition using CNN-based algorithm on high-speed camera imaging: Framework for fast and robust high-speed camera object recognition based on population data cleansing and data ensemble. In *Proceedings of the IEEE International Conference on Pattern Recognition (ICPR)*, pages 2025–2032.

Narang, M., Rana, M., Patel, J., D'Souza, S., Onyechie, P., Berry, S., Tayefeh, M., and Barari, A. (2021). Fighting COVID: An autonomous indoor cleaning robot (AICR) supported by artificial intelligence and vision for dynamic air disinfection. In *Proceedings of the IEEE International Conference on Industry Applications*, pages 1146–1153.

Nasralla, M. M., Rehman, I. U., Sobnath, D., and Paiva, S. (2019). Computer vision and deep learning-enabled UAVs: Proposed use cases for visually impaired people in a smart city. In *Proceedings of IAPR International Conference on Computer Analysis of Images and Patterns (CAIP). LNCS 9256, Part I*, pages 91–99.

Nowozin, S., Rother, C., Bagon, S., Sharp, T., Yao, B., and Kohli, P. (2011). Decision tree fields. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1668–1675.

Olivares-Alarcos, A., Bessler, D., Khamis, A., Goncalves, P., Habib, M., Bermejo-Alonso, J., Barreto, M., Diab, M., Rosell, J., Quintas, J., Olszewska, J. I., Nakawala, H., Pignaton de Freitas, E., Gyrard, A., Borgo, S., Alenya, G., Beetz, M., and Li, H. (2019). A review and comparison of ontology-based approaches to robot autonomy. *The Knowledge Engineering Review*, 34:1–38.

Olszewska, J. I. (2013). Semantic, automatic image annotation based on multi-layered active contours and decision trees. *International Journal of Advanced Computer Science and Applications*, 4(8):201–208.

Olszewska, J. I. (2015a). Active contour based optical character recognition for automated scene understanding. *Neurocomputing*, 161(C):65–71.

Olszewska, J. I. (2015b). "Where Is My Cup?" – Fully automatic detection and recognition of textureless objects in real-world images. In *Proceedings of IAPR International Conference on Computer Analysis of Images and Patterns (CAIP). LNCS 9256, Part I*, pages 501–512.

Olszewska, J. I. (2017). Active-contours based-on face emotion patterns. In *Proceedings of IAPR International Conference on Computer Analysis of Images and Patterns (CAIP) Workshop*, pages 59–70.

Olszewska, J. I. (2018). Discover intelligent vision softwares. In *DDD Scotland*.

Olszewska, J. I. (2019a). D7-R4: Software development life-cycle for intelligent vision systems. In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development (KEOD)*, pages 435–441.

Olszewska, J. I. (2019b). Designing transparent and autonomous intelligent vision systems. In *Proceedings of the International Conference on Agents and Artificial Intelligence (ICAART)*, pages 850–856.

Olszewska, J. I. (2020). Developing intelligent vision software and the future of AI. In *ACM Future Worlds Symposium (FWS)*.

Olszewska, J. I. (2021). Algorithms for intelligent vision systems. In *Canadian Mathematical Society 75th Anniversary Summer Meeting (CMS)*.

Olszewska, J. I. and McCluskey, T. L. (2011). Ontology-coupled active contours for dynamic video scene understanding. In *Proceedings of the IEEE International Conference on Intelligent Engineering Systems*, pages 369–374.

Pignaton de Freitas, E., Olszewska, J. I., Carbonera, J. L., Fiorini, S., Khamis, A., Sampath Kumar, V. R., Barreto, M., Prestes, E., Habib, M., Redfield, S., Chibani, A., Goncalves, P., Bermejo-Alonso, J., Sanz, R., Tosello, E., Olivares Alarcos, A., Konzen, A. A., Quintas, J., and Li, H. (2020). Ontological concepts for information sharing in cloud robotics. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–14.

Prakash, A., Ramakrishnan, N., Garg, K., and Srikanthan, T. (2020). Accelerating computer vision algorithms on heterogeneous edge computing platforms. In *Proceedings of the IEEE Workshop on Signal Processing Systems*, pages 1–6.

Prestes, E., Houghtaling, M., Goncalves, P., Fabiano, N., Ulgen, O., Fiorini, S. R., Murahwi, Z., Olszewska, J. I., and Haidegger, T. (2021). IEEE 7007: The first global ontological standard for ethical robotics and automation systems. *IEEE Robotics and Automation Magazine*, 28(4):120–124.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., and Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, pages 33–44.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, pages 91–99.

Ryabchikov, I., Teslya, N., and Druzhinin, N. (2020). Integrating computer vision technologies for smart surveillance purpose. In *Proceedings of the IEEE Conference of Open Innovations Association (FRUCT)*, pages 392–401.

Samani, E. U., Yang, X., and Banerjee, A. G. (2021). Visual object recognition in indoor environments using topologically persistent features. *IEEE Robotics and Automation Letters*, 6(4):7509–7516.

Shirazi, M. S., Patooghy, A., Shisheie, R., and Haque, M. M. (2020). Application of unmanned aerial vehicles in smart cities using computer vision techniques. In *Proceedings of the IEEE International Smart Cities Conference*, pages 1–7.

Sui, X. (2021). Research on object location method of inspection robot based on machine vision. In *Proceedings of the IEEE International Conference on Intelligent Transportation, Big Data and Smart City*, pages 724–727.

Wang, J., Sun, K., Cheng, t., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., and Xiao, B. (2021). Deep high-resolution representation learning for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3349–3364.

Wang, J. K., Yan, F., Aker, A., and Gaizauskas, R. (2014). A Poodle or a Dog? Evaluating automatic image annotation using human descriptions at different levels of granularity. In *Proceedings of the International Conference on Computational Linguistics*, pages 38–45.

Wang, Z., Yan, X., Han, Y., and Sun, M. (2019). Ranking video salient object detection. In *Proceedings of the ACM International Conference on Multimedia (MM)*, pages 873–881.

Wilding, S., Walker, P., Clinton, S., Williams, D., and Olszewska, J. I. (2020). Safe human-computer interface based on an efficient image processing algorithm. In *Proceedings of the IEEE International Symposium on Computational Intelligence and Informatics (CINTI)*, pages 65–70.

Winfield, A. F. T., Booth, S., Dennis, L. A., Egawa, T., Hastie, H., Jacobs, N., Muttram, R., Olszewska, J. I., Rajabiyazdi, F., Theodorou, A., Underwood, M., Wortham, R. H., and Watson, E. (2021). IEEE P7001: A proposed standard on transparency. *Frontiers on Robotics and AI*, 8:1–16.

Yezzi, A., Sundaramoorthi, G., and Benyamin, M. (2019). PDE acceleration for active contours. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12310–12320.

Zeng, H., Song, X., Chen, G., and Jiang, S. (2020). Learning scene attribute for scene recognition. *IEEE Transactions on Multimedia*, 22(6):1519–1530.

Zhang, D., Islam, M. M., and Lu, G. (2012). A review on automatic image annotation techniques. *Pattern Recognition*, 45(1):346–362.

Zhang, T., Li, Q., Zhang, C.-S., Liang, H.-W., Li, P., Wang, T.-M., Li, S., Zhu, Y.-L., and Wu, C. (2017). Current trends in the development of intelligent unmanned autonomous systems. *Frontiers of Information Technology and Electronic Engineering*, 18(1):68–85.