

A Systematic Mapping on Detection of Human Mouth Landmarks

Pedro Henrique D’Almeida G. Rissato
Prog. de PG em Computação Aplicada
DCM-FFCLRP
Universidade de São Paulo
Ribeirão Preto, São Paulo, Brasil
pedrorissato@usp.br

Renato de Freitas Bulcão-Neto
Instituto de Informática
Universidade Federal de Goiás
Goiânia, Goiás, Brasil
rbulcao@ufg.br

Alessandra Alaniz Macedo
Prog. de PG em Computação Aplicada
DCM-FFCLRP
Universidade de São Paulo
Ribeirão Preto, São Paulo, Brasil
ale.alaniz@usp.br

Abstract—Facial landmarks represent regions of interest whose detection and localization generate features supporting the identification of movements, feelings, and reactions. Most facial feature detection algorithms focus on entire semantic areas, such as the region of a mouth which allows grained manipulation that is essential for a wide domain variety. This paper describes a systematic mapping of the detection of landmarks in human faces and their application domains. The identification and selection methods of primary studies include automatic search on information sources, inclusion, and exclusion criteria over 344 scientific papers from 2015 and 2021, from which we analyzed and synthesized 115 primary studies. Our analysis considered the implementation of methods, types, and uses of data extracted from the mouth. The mapping brought exciting information as new methods, datasets, and domains researched through the time interval reviewed as well research gaps that can be explored.

Index Terms—systematic mapping, human face, detection and recognition of landmark

I. INTRODUCTION

In the realm of computer vision, a landmark is a recognizable natural or artificial feature used for navigation. On the other hand, facial landmarks are regions of interest on the face such as the nose tip, eyes corners, chin, mouth corners, nostril corners, eyebrow arcs, and ear lobes. The detection and the localization of facial landmarks generate features supporting the identification of movements, feelings, and reactions. For example, information extracted from landmarks identifies expressions [1], [2] and performs face recognition [3], [4] such as a tired person driving a car [5]–[7], a person smiling [8], [9], a person paying attention to a video call [10], [11] or even a person having a stroke [12], [13]. Therefore, landmarks play a prominent role in the extraction of information from regions of human faces [14].

For analysis purposes, most landmark detection algorithm focuses on an entire facial semantic region, such as the whole region of a mouth. Mouth state detection empowers distinct applications such as the detection of a smile (a type of feeling), the diagnosis of health conditions (stroke and others), the

start of a recording of speeches, and the correct training of phonological exercises.

This paper presents a Systematic Mapping (SM) study [15], [16] that charts primary research on detection and recognition of mouth landmarks in human faces. The SM protocol includes automatic search on six relevant information sources (ACM, Engineering Village, IEEE, ScienceDirect, Scopus, and Web of Science), and inclusion/exclusion criteria over 344 papers published between 2015 and 2021, inclusive. From 344 articles, we analyzed and synthesized 115 primary studies by considering the implementation of methods and the types and the uses of data extracted from the mouth.

This SM intends to raise all algorithms, feature extraction techniques, image/video database and domains in which landmarks are applied on the human face, especially involving the mouth, to serve as a guide on the subject, being reproducible over time, as well as exploring possible research gaps.

This paper’s structure is as follows: Section 2 describes related works among the field; Section 3 describes the SM protocol; Section 4 reports the data extraction and synthesis of the relevant studies, and Section 5 discusses the findings and Section 6 brings final remarks.

II. RELATED WORK

In [17], the authors established 5 stages for defining landmarks: (i) objective definition, (ii) dataset selection, (iii) regions of interest, (iv) model definition and, (v) model training and evaluation, along with eleven datasets (Multi-Pie, MUCT, XM2VTS, Menpo benchmark, 300W, HELEN, AFLW, LFW, BioID, PUT and Caltech 10k web faces), a comparative analysis between OpenCV and Dlib, concluded that Histogram of Gradient (HOG) algorithm applied to Dlib outperforms OpenCV and Viola-Jones algorithm in most scenarios.

The work in [18] reviewed facial landmark detection in the realms of Deep Learning. Landmark extraction was separated into five categories: (i) sparse facial landmarks detection; (ii) dense facial landmarks detection; (iii) landmarks detection with Recurrent Neural Networks (RNN); (iv) landmark detection with likelihood maps, and (v) landmark detection with multi-task learning. In landmark tracking RNN, RED-Net, Long Short-Term Memory (LSTM), Cascade Regression,

The authors would like to thank FAPESP (16/13206-4), CNPq (302031/2016-2), and CAPES/FAPEG (88887.305511/201800) for supporting this work. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

and Convolutional Neural Networks were the most chosen methods. The authors came to a conclusion that Deep Learning methods outperform others, despite efficiency issues and complexity, especially on real-time video analysis.

An extensive approach was performed by [19] where facial landmarks were categorized into holistic methods, Constrained Local Model (CLM) methods, and the regression-based methods. The authors enumerate 16 most used databases for benchmarking as well as 23 softwares for landmark purposes (18 for academic use and five for commercial use). In short, the authors concluded that there is no single algorithm that can approach all cases and one database that contains all variations for a unique model generalized. In addition, landmark detection relies on face detection which may have a high computation cost for real-time processing. Finally, the authors conclude that an approach that covers all three methods described (holistic, CLM and regression-based) must benefit new generalized models capable of covering a wide variety of "in-the-wild" conditions.

The articles cited, despite bringing a valuable contribution to the area, did not establish a period of time of the studies, so little demonstrated how the survey was carried out, and, in general, approach the theme only about the vision of neural networks. Our work is also aimed at establishing a protocol that is reproducible over time, and, presenting all studies, in the last seven years, which dealt with landmarks theme on the human face, mainly involving the mouth, so that it is possible to evaluate which techniques, algorithms, and datasets are best applicable to a particular domain, considering that, as described above by the other surveys, there is no single technique, nor database that can encompass all scenarios.

III. THE SYSTEMATIC MAPPING PROTOCOL

Systematic Mapping (SM) consists of the identification and classification of primary studies of interest. Moreover, SM supports gathering information about a subject, to identify best and regular practices [16].

A. Research Questions and Search Strategy

To achieve the goal, we conducted steps that included the planning of SM, the execution of the SM protocol, and publishing of results. In the planning stage, we defined research questions and a research protocol. In SM, research questions tend to be more generic than a Systematic Review [20], since SM establishes the subject trends or state-of-the-art.

Based on the goal of identifying and analyzing the use of mouth landmarks in human faces, we formulated the following Research Questions (RQs):

- RQ1: Which do primary studies apply the landmark technique for image pattern recognition in human faces?
- RQ2: Which are the most used techniques to extract or to select features?
- RQ3: Which are the algorithms used to recognize the patterns by considering landmarks?
- RQ4: Which are the datasets used to perform the analysis?

RQ5: What are the domains in which the technique involving landmarks in the human mouth acts?

The question structure suggests four points of views [21] called PICO (Population, Intervention, Comparison, and Outcome) as Table I describes:

TABLE I
CRITERIAS FOR VIEWPOINT IN SOFTWARE ENGINEERING REVIEW
PROCESS

Criteria	Answered by research question #
Population	RQ1
Intervention	RQ2
Comparison	RQ3, RQ4 and RQ5
Outcomes	Information relationship between RQ3, RQ4 and RQ5

As seen in Table I, the RQs consist of information about (i) which will be observed in the context of the systematic mapping (Intervention: RQ2), (ii) techniques for feature extraction (Intervention: RQ2), (iii) who will be noticed (Population: RQ1), and (iv) expected results (Results: RQ3 - algorithms, RQ4 - datasets and RQ5 - domains) and their relationship intersections.

The RQ2 aims to identify techniques often used to manipulate features among pattern recognition processes. RQ3 investigates the recent use of landmarks by pattern recognition algorithms. RQ4 aims to discover datasets used in human face pattern researches. Finally, RQ5 identifies which domain fields are explored by the studies.

To perform automatic searches, the next step was the formulation of a search string. The first search string was the following keywords linked by *AND* logic operators: mouth, landmark, and detection. For each keyword, we defined synonyms using the *OR* operator as described:

(mouth AND landmark AND detection) OR (mouths AND landmarks AND detections) OR (mouth AND face AND landmark AND detection)

After this naïve search string performed, the search string exploited the keywords returned by some of the papers of interest. We repeated this method until we achieved the last search string by acquiring a large number of papers as minimizing unrelated ones, which was:

facial AND mouth AND landmark AND detection

Next, we looked for primary studies on scientific information sources by using the search engine. We chose the following sources: ACM Digital Library, Engineering Village, IEEE Xplore Digital Library, ScienceDirect, Scopus, and Web of Science. The criteria to choose them include the relevance for the Computer Science area, the availability of a search mechanism on the Web with logical expression support, the possibility of search over papers metadata, and a presence of a temporal filtering mechanism. We also defined the selection of solely English-written articles published from 2015 to 2021, inclusive.

We submitted the final query on 2021-09-21 to the selected digital repositories. Then we eliminated duplicated papers,

keeping the unique ones into a digital repository with most results; next, the resulting papers were analyzed using inclusion and exclusion criteria, as we describe next.

B. Selection of Primary Studies

The definition and the assignation of inclusion and exclusion criteria based on RQs support the selection of studies. Inclusion criteria are features that the retrieved studies must have to be included in the resulting mapping, while exclusion criteria are features that disqualify studies from the inclusion.

When a retrieved study matches with at least one exclusion criterion, we excluded this study from the mapping. For the systematic mapping, we defined the following Exclusion Criteria (EC):

- EC1: It is not a primary study.
- EC2: The study is not a full paper.
- EC3: The study is an older version of another study.
- EC4: The study is older than 7 (seven) years considering 2015-2021.
- EC5: The study does not employ the landmark technique on human face image pattern recognition.
- EC6: The study does not have an abstract.
- EC7: The full paper is not available on the Web or the *Portal de Periódicos da Capes*¹.
- EC8: It is not an English written study.
- EC9: The study does not make use of the landmark technique on the mouth, even indirectly.
- EC10: It is a duplicate study (an identical study found on another database).

After selection by exclusion criteria, we analyzed the other studies according to Inclusion Criterion (IC). If at least one of the requirements was met, we carried out an in-depth reading of the study.

- IC1: The study presents a comparative research of landmark techniques for pattern recognition in images of human faces.
- IC2: The study presents a new approach to landmark techniques for pattern recognition in images of human faces.
- IC3: The study presents a landmark technique for pattern recognition in images of human faces.

Initially, 344 papers were retrieved from the databases considering papers from 2015 to 2021 inclusive in which 203 duplicate studies were excluded. Next, we read the title, summary, and keywords of each of the 141 remaining studies applying the exclusion and inclusion criteria. As a result, we selected 137 potentially relevant studies based only on the reading and interpretation of the content of the respective metadata.

In the data extraction activity, described in the next section, we utterly read each one of the 135 remaining studies.

¹Available at <https://www.periodicos.capes.gov.br>.

IV. DATA EXTRACTION AND SYNTHESIS

The 137 primary studies remained from the study selection stage were utterly read in data extraction phase, and 22 papers were excluded by exclusion criteria (Table II). As a result, 115² papers were selected to extract the answers for the RQs (data synthesis).

TABLE II
NUMBER OF PAPERS EXCLUDED BY EXCLUSION CRITERIA AND SYSTEMATIC MAPPING STAGE. NO STUDIES WERE REMOVED BY EXCLUSION CRITERIAS EC2, EC4, EC6, EC8 AND, EC9

Stage	EC1	EC3	EC5	EC7	EC9	EC10	TOTAL
Studies Selection	4	0	0	0	0	203	207
Data Extraction	0	1	3	17	1	0	22
TOTAL	4	1	3	17	1	203	229

The contributions of each information source after reading and excluding papers based on criteria of this SM are Scopus (82), IEEE Digital Library (19), Engineering Village (10), Web of Science (3) and, ACM Digital Library (1). All five papers from ScienceDirect were eliminated by being duplicated into another information source (EC10). Figure 1 illustrates the number os papers published by year and information source.

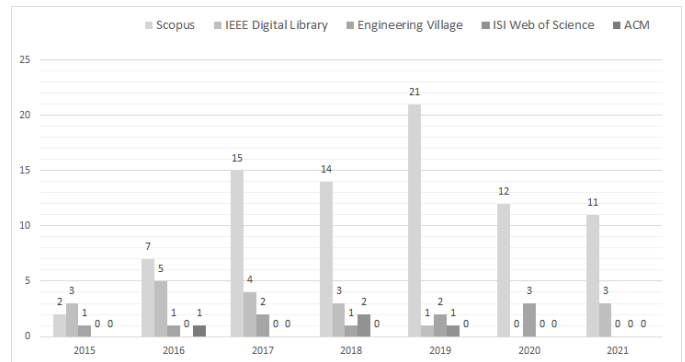


Fig. 1. Distribution of studies along years by information source.

A. Landmark with Features Extraction and Selection Techniques

To answer the RQ2, 115 different studies about landmarks treat of this theme. Precisely, 75 proposed new approaches, 23 of these studies exploited the HOG [22], 11 studies applied Eye Aspect Rate (EAR) [23] and SVM Support Vector Machine [24], and nine studies described Mouth Aspect Ratio (MAR) [25].

The new approaches are proposed by scientists who developed their methodology to extract and select features of face images. The other most used techniques are well-founded techniques with excellent results published in the literature for a long time [1], [26]–[38], including the patented SIFT technique [7], [8], [31], [34], [38], [39].

²All papers found and their relationship with RQs are available at <https://drive.google.com/file/d/1Ehz7iIj8Bd-LA9SRcPNTMKYypQd9mdMB/view?usp=sharing>.

Although, most studies applied their concept technique to extract features, HOG plays a significant role in landmark field, usually combined with SVM for feature selection that most describes the domain they are at [1], [30], [37], [40]–[57]. Lastly, EAR and MAR usage derives from the most research domain: Driver Fatigue/Inebriation/Drowsiness, as these techniques are used for verifying closed eyes or yawning on a driver [42], [58]–[60].

B. Landmarks with Pattern Recognition Algorithm

The answers to the RQ3 describe the majority of studies (52) develop its learning algorithm to recognize a pattern. The second most used algorithm is (26) Support Vector Machine [24]. (18) Convolutional Neural Networks (CNN) [61] were in the third position. The fourth is the (11) Viola-Jones algorithm [62] followed by (7) K-nearest neighbors (K-NN) [63].

As the majority of studies have created new algorithms to recognize mouth landmarks, it suggests an open research problem. There is no consensus about which technique provides the most acceptable solution and highlights the growth of this field.

However, it is noteworthy that SVM appears to be more connected with studies relative to driver fatigue/inebriation/drowsiness. CNN is applied to the same field as well, tied with defining landmarks, considering neural networks have a great precision combined with recognition/detection speed, although demanding high computational power to provide these qualities.

C. Datasets

Most studies created their datasets. (41) 300 Faces In-The-Wild Challenge [64] was the second most used dataset followed by the (14) HELEN dataset [65]. (11) Labeled Face Parts in the Wild [66] occupies the third position followed by (8) Annotated Faces in The Wild (AFW) [67] in fourth position.

The 300-W is composed of a total of 600 outdoor and indoor images from sites. This database contains one to seven faces in variations of positions, expressions, and light conditions. A tool developed by [64] automatically arranges the landmarks.

Consultants created the HELEN dataset and manually positioned the landmarks using Flickr images of faces with a resolution higher than 500 pixels in various light conditions, poses, and facial expressions.

LFPW contains only images of human faces acquired from sites by using simple queries. It added landmarks manually. The Annotated Faces in the Wild (AFW) comprises random Flickr images obtained with variations of background, face positions, expressions, and appearance (age, sunglasses, makeup, skin color, etc.). The 300-W allows the inclusion of landmarks in face images by automated methods. Despite that, the LFPW and HELEN datasets detect parts on the human face (eyes, mouth, nose, etc.). Regarding the Annotated Faces in the Wild aimed at the detection of the human face.

D. Applications of Studies Found

Most of the studies try to determine driver fatigue, inebriation, or drowsiness (30). Define landmarks was the second most research theme (25), followed by facial expression recognition (20). Head pose estimation, which is used to define landmarks and to improve accuracy on face detection (8). The fifth domain was facial recognition (4), facial construction/reconstruction (3), stroke recognition, face detection and, visual voice detection were in the sixth position with three studies each.

Domains like attention detection, estimate age, fatigue detection, gender recognition, micro-facial expressions recognition, mouth tracking, and smile detection have endeavored two times each. Finally, domains such as classify face parts, classify physical exercise intensity, detect face with occlusion, detect facial paralysis, detecting swearing on sports events, driver distraction, estimate heartbeat, track eyes, generate a 3D model, grading facial paralysis, detect human face contour, perform a landmark semantic segmentation, localize inner eye canthus, provide a markless motion capture and prevent cheating in online tests were cited one time by studies found.

V. DISCUSSION

This paper achieved its goals by mapping 115 studies, from 2015 to 2021 inclusive, in the field of computer vision and computer science, supported by a string defined to extract the necessary information and answer five research questions. Our results help to determine the actual state-of-the-art about human face pattern recognition, especially evolving the mouth.

The bubble graph in Figure 2 unifies the answers of RQ3, RQ4, and RQ5 after analyzing the 115 articles resulting from the systematic mapping process³. Figure 2 brought the five topics that most appeared in SM, considering the areas of algorithms, database, and domains, concatenated in a three-dimensional chart, where it is possible to visualize the relationships between them, as well as possible research gaps.

It is noteworthy that 16 of 29 driver fatigue/inebriation/drowsiness citations had their own dataset built due to the lack of a specific dataset for the specified domain [68], [69]. However, considering the number of research papers that explored this domain, their datasets built, and other datasets used for building, a new dataset unifying those could be beneficial for new models for this domain problem.

There is a homogeneous distribution between the Viola-Jones and Convolutional Neural Networks. The Viola-Jones algorithm is the precursor of human facial recognition in images with very diverse datasets of frontal face images by considering different lighting, pose, and occlusion conditions. Consequently, they are capable of producing robust implementations of generalized models of recognition facial. Convolutional Neural Networks are becoming more used as

³It is that a single article can simultaneously address two domains and several algorithms. So Figure 2 represents the sum of these citations and not their uniqueness in the form of an essay.

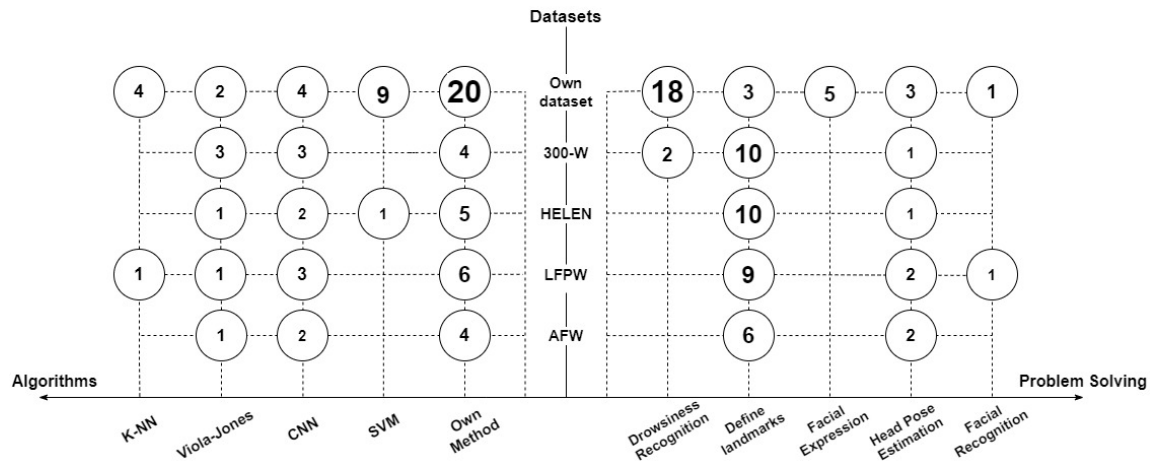


Fig. 2. Mapping among the first five algorithms, datasets, and problem domains.

computational power arises and transfer learning [70]–[72] techniques are applied.

The domain of facial expression recognition was addressed by 20 studies, using a total of 17 unique datasets; of these, only five appeared among the most used databases (Own Dataset), the second most cited being the Extended Cohn-Kanade Dataset (CK+) with seven citations. Despite the expressiveness of recognition of facial expressions, there is no consensus about datasets for building generalized models in the area. There is also an opportunity for research on comparison or even concatenation of such datasets to build more robust facial expression recognition models.

VI. FINAL REMARKS

In this paper, we synthesized the literature involving landmark detection and recognition in the human face, especially using mouth region. By answering five research questions, we found the algorithms, datasets, feature selections, and field domains. As a result, we demonstrated new research opportunities to build models and applications with human face interactions using landmarks.

REFERENCES

- [1] K. Lekdioui, R. Messoussi, Y. Ruichek, Y. Chaabi, and R. Touahni, "Facial decomposition for expression recognition using texture/shape descriptors and svm classifier," *Signal Processing: Image Communication*, vol. 58, pp. 300–312, 2017.
- [2] S. Anwar, M. Milanova, A. Bigazzi, L. Bocchi, and A. Guazzini, "Real time intention recognition," in *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, 10 2016, pp. 1021–1024.
- [3] A. Juhong and C. Pintavirooj, "Face recognition based on facial landmark detection," vol. 2017-January, Hokkaido, Japan, 2017, pp. 1–4.
- [4] L. Wang, H. Zhang, and Z. Wang, in *Component based representation for face recognition*, vol. 3, Qingdao, China, 2015, pp. 1275–1278.
- [5] B. Reddy, Y. Kim, S. Yun, C. Seo, and J. Jang, "Real-time driver drowsiness detection for embedded system using model compression of deep neural networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 7 2017, pp. 438–445.
- [6] L. Zhao, Z. Wang, X. Wang, and Q. Liu, "Driver drowsiness detection using facial dynamic fusion information and a dbn," *IET Intelligent Transport Systems*, vol. 12, no. 2, pp. 127–133, 2018.
- [7] R. Zheng, C. Tian, H. Li, M. Li, and W. Wei, "Fatigue detection based on fast facial feature analysis," vol. 9315, Gwangju, Korea, Republic of, 2015, pp. 477–487.
- [8] D. Cui, G.-B. Huang, and T. Liu, "Elm based smile detection using distance vector," *Pattern Recognition*, vol. 79, pp. 356–369, 2018.
- [9] P. Wu, H. Liu, C. Xu, Y. Gao, Z. Li, and X. Zhang, "How do you smile? towards a comprehensive smile analysis system," *Neurocomputing*, vol. 235, pp. 245–254, 2017.
- [10] X. Min, G. Zhai, and K. Gu, "Visual attention on human face," in *2015 Visual Communications and Image Processing (VCIP)*, 12 2015, pp. 1–4.
- [11] Y. Ren, Z. Wang, M. Xu, H. Dong, and S. Li, in *Learning Dynamic GMM for Attention Distribution on Single-Face Videos*, vol. 2017-July, Honolulu, HI, United states, 2017, pp. 1632–1641.
- [12] O.-M. Foong, K.-W. Hong, and S.-P. Yong, "Droopy mouth detection model in stroke warning," Kuala Lumpur, Malaysia, 2016, pp. 616–621.
- [13] Y. Zhuang, O. Uribe, M. Mcdonald, I. Lin, D. Arteaga, W. Dalrymple, B. Worrall, A. Southerland, and G. Rohde, "Pathological facial weakness detection using computational image analysis," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 4 2018, pp. 261–264.
- [14] R. Angulu, A. O. Adewumi, and J.-R. Tapamo, "Landmark localization approach for facial computing," Umhlanga, Durban, South africa, 2017.
- [15] K. R. Felizardo, E. Y. Nakagawa, S. C. P. F. Fabri, and F. C. Ferrari, *Revisão Sistemática da Literatura em Engenharia de Software*, 1st ed. Elsevier, 2017.
- [16] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," in *EASE*, vol. 8, 2008, pp. 68–77.
- [17] B. Johnston and P. Chazal, "A review of image-based automatic facial landmark identification techniques," *EURASIP Journal on Image and Video Processing*, vol. 2018, p. 86, 09 2018.
- [18] M. Bodini, "A review of facial landmark extraction in 2d images and videos using deep learning," *Big Data and Cognitive Computing*, vol. 3, no. 1, 2019.
- [19] Y. Wu and Q. Ji, "Facial landmark detection: A literature survey," *Int. J. Comput. Vision*, vol. 127, no. 2, p. 115–142, Feb. 2019.
- [20] S. C. P. F. Fabbri, K. R. Felizardo, F. C. Ferrari, E. C. M. Hernandez, F. R. Octaviano, E. Y. Nakagawa, and J. C. Maldonado, "Externalising tacit knowledge of the systematic review process," *IET Software*, vol. 7, no. 6, pp. 298–307, 2013.
- [21] K. BA and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," vol. 2, 01 2007.
- [22] W. T. Freeman and M. Roth, "Orientation histograms for hand gesture recognition," MERL - Mitsubishi Electric Research Laboratories, Cambridge, MA 02139, Tech. Rep. TR94-03, 12 1994.
- [23] T. Soukupová and J. Cech, "Real-time eye blink detection using facial landmarks," 2016.
- [24] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 9 1995.

- [25] P. Awasekar, M. Ravi, S. Doke, and Z. Shaikh, "Driver fatigue detection and alert system using non-intrusive eye and yawn detection," *International Journal of Computer Applications*, vol. 180, pp. 1–5, 05 2018.
- [26] F. Becerra-Riera, H. Méndez-Vázquez, A. Morales-Gonzalez, and M. Tistarelli, "Age and gender classification using local appearance descriptors from facial components," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, 10 2017, pp. 799–804.
- [27] C.-Y. Chang, M.-J. Cheng, and M. H.-M. Ma, "Application of machine learning for facial stroke detection," vol. 2018-November, Shanghai, China, 2018.
- [28] S. G. Shivashankar and S. Hiremath, "Emotion sensing using facial recognition," Bengaluru, India, 2017, pp. 830–833.
- [29] Z. H. Aung and P. Ritthipravat, in *Robust visual voice activity detection using Long Short-Term Memory recurrent neural network*, vol. 9431, Auckland, New Zealand, 2016, pp. 380–391.
- [30] S. Afshar and A. A. Salah, "Facial expression recognition in the wild using improved dense trajectories and fisher vector encoding," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 06 2016, pp. 1517–1525.
- [31] Z. Xie, Y. Jin, P. Bian, and W. Zhou, "Facial landmark detection via self-adaptation model and multi-task feature learning," vol. 0, Chengdu, China, 2016, pp. 113–117.
- [32] W. Zhang, W. Wang, S. Zhao, and B. Sun, "Gray-edge-hog feature based cascaded learning for facial landmark detection," vol. 189, Beijing, China, 2018.
- [33] A. Kumar and R. Patra, in *Driver drowsiness monitoring system using visual behaviour and machine learning*, Penang Island, Malaysia, 2018, pp. 339–344.
- [34] P. Bian, Z. Xie, and Y. Jin, "Multi-task feature learning-based improved supervised descent method for facial landmark detection," *Signal, Image and Video Processing*, vol. 12, no. 1, pp. 17–24, 2018.
- [35] I. Gupta, N. Garg, A. Aggarwal, N. Nepalia, and B. Verma, "Real-time driver's drowsiness monitoring based on dynamically varying threshold," Noida, India, 2018.
- [36] Y. Zhang, F. Jiang, and R. Shen, in *Region-Based Face Alignment with Convolution Neural Network Cascade*, vol. 10636 LNCS, Guangzhou, China, 2017, pp. 300–309.
- [37] X. Chang and W. Skarbek, "Facial expressions recognition by animated motion of candid 3d model," vol. 10808, Wilga, Poland, 2018, pp. ARIES – Accelerator Research and Innovation for European Science and Society (CERN, EU H2020); Committee of Electronics and Telecommunications, Polish Academy of Sciences; EuroFusion Collaboration; EuroFusion Poland; PKOpto – Polish Committee of Optoelectronics of SEP–The Association of Polish Electrical Engineers; PSP – Photonics Society of Poland.
- [38] B. S. Riggan, N. J. Short, and S. Hu, in *Thermal to Visible Synthesis of Face Images Using Multiple Regions*, vol. 2018-January, Lake Tahoe, NV, United States, 2018, pp. 30–38.
- [39] K. Lekdioui, Y. Ruichek, R. Messoussi, Y. Chaabi, and R. Touahni, "Facial expression recognition using face-regions," Fez, Morocco, 2017, p. CNRST; et al.; Faculty of Medicine and Pharmacy of Fez; Faculty of Sciences of Fez; TICSM; University of Sidi Mohamed Ben Abdellah of Fez.
- [40] C.-Y. Chang, M.-J. Cheng, and M.-M. Ma, "Application of machine learning for facial stroke detection," vol. 2018-November, 2019.
- [41] V. Guttha, H. Kondakindi, and V. Bhatti, "Automated feedback generation system using facial emotion recognition," 2018, pp. 975–981.
- [42] S. Mohanty, S. Hegde, S. Prasad, and J. Manikandan, "Design of real-time drowsiness detection system using dlib," 2019.
- [43] A. Kumar and R. Patra, "Driver drowsiness monitoring system using visual behaviour and machine learning," 2018, pp. 339–344.
- [44] V. Venkata Sai Vardhan, N. Ritish Kumar Reddy, K. Jaya Surya, J. Uday Kiran, and A. Kumar, "Driver's drowsiness detection based on facial multi-feature fusion," vol. 1998, no. 1, 2021.
- [45] S. Shivashankar and S. Hiremath, "Emotion sensing using facial recognition," 2018, pp. 830–833.
- [46] J. Josephine Julina and T. Sharmila, "Facial emotion recognition in videos using hog and lbp," 2019, pp. 56–60.
- [47] Z. Xie, Y. Jin, P. Bian, and W. Zhou, "Facial landmark detection via self-adaptation model and multi-task feature learning," vol. 0, 2016, pp. 113–117.
- [48] M. Santosh and A. Sharma, "Fusion of multi representation and multi descriptors for facial expression recognition," vol. 1057, no. 1, 2021.
- [49] W. Zhang, W. Wang, S. Zhao, and B. Sun, "Gray-edge-hog feature based cascaded learning for facial landmark detection," vol. 189, 2018.
- [50] Y. Fei, B. Li, H. Wang, and L. Tian, "Long short-term memory network based fatigue detection with sequential mouth feature," 2020, pp. 217–222.
- [51] P. Bian, Z. Xie, and Y. Jin, "Multi-task feature learning-based improved supervised descent method for facial landmark detection," *Signal, Image and Video Processing*, vol. 12, no. 1, pp. 17–24, 2018.
- [52] S. Dey, S. Chowdhury, S. Sultana, M. Hossain, M. Dey, and S. Das, "Real time driver fatigue detection based on facial behaviour along with machine learning approaches," 2019, pp. 135–140.
- [53] S. Dey, M. Islam, S. Chowdhury, M. Islam, M. Ali Hossain, and S. Das, "Real time tracking of driver fatigue and inebriation maintaining a strict driving schedule," 2019, pp. 9–14.
- [54] A. M. Al-madani, A. T. Gaikwad, V. Mahale, Z. A. Ahmed, and A. A. Shareef, "Real-time driver drowsiness detection based on eye movement and yawning using facial landmark," in *2021 International Conference on Computer Communication and Informatics (ICCCI)*, Jan 2021, pp. 1–4.
- [55] I. Gupta, N. Garg, A. Aggarwal, N. Nepalia, and B. Verma, "Real-time driver's drowsiness monitoring based on dynamically varying threshold," 2018.
- [56] T. Zhang, Z. Chen, and C. Ouyang, "Research on driver fatigue detection," vol. 10609, 2017.
- [57] Z. Aung and P. Ritthipravat, "Robust visual voice activity detection using long short-term memory recurrent neural network," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9431, pp. 380–391, 2016.
- [58] S. Mehta, P. Mishra, A. Bhatt, and P. Agarwal, "Ad3s: Advanced driver drowsiness detection system using machine learning," vol. 2019-November, 2019, pp. 108–113.
- [59] A.-C. Phan, N.-H.-Q. Nguyen, T.-N. Trieu, and T.-C. Phan, "An efficient approach for detecting driver drowsiness based on deep learning," *Applied Sciences (Switzerland)*, vol. 11, no. 18, 2021.
- [60] P. Inthanon and S. Mungsing, "Detection of drowsiness from facial images in real-time video media using nvidia jetson nano," 2020, pp. 246–249.
- [61] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1–6.
- [62] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, 12 2001, pp. I–I.
- [63] E. Fix and J. L. Hodges, "Discriminatory analysis. nonparametric discrimination: Consistency properties," *International Statistical Review / Revue Internationale de Statistique*, vol. 57, no. 3, pp. 238–247, 1989.
- [64] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: database and results," *Image and Vision Computing*, vol. 47, 01 2016.
- [65] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, "Interactive facial feature localization," in *Proceedings of the 12th European Conference on Computer Vision - Volume Part III*, ser. ECCV'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 679–692.
- [66] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars," in *CVPR 2011*, 2011, pp. 545–552.
- [67] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2879–2886.
- [68] A. Atla, R. Tada, V. Sheng, and N. Singireddy, "Sensitivity of different machine learning algorithms to noise," *J. Comput. Sci. Coll.*, vol. 26, no. 5, p. 96–103, May 2011.
- [69] V. Vapnik, *The Support Vector Method of Function Estimation*. Boston, MA: Springer US, 1998, pp. 55–85.
- [70] S. Bozinovski, "Reminder of the first paper on transfer learning in neural networks, 1976," *Informatica*, vol. 44, no. 3, Sep. 2020.
- [71] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI global, 2010, pp. 242–264.
- [72] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.