УДК 004.03:681.5

<https://doi.org/10.17721/1812-5409.2020/1-2.15>

Пашко А.О., д.ф.-м.н., проф
Тесленко О.В., аспірант

A. O. Paschko, D,
O. V. Teslenko, PhD stud.

Огляд основних підходів до визначення активності водія з камер встановлених в автомобілі

A survey of driver activity recognition from cameras installed in a car

Київський національний університет імені
Тараса Шевченка, факультет комп'ютерних наук
і кібернетики., 01601, м. Київ, вул.
Володимирська 64/13,
e-mail: aap2011@ukr.net
e-mail: oteslenkoua@gmail.com,

Taras Shevchenko National University of Kyiv,
01601, Kyiv, Volodymyrska st. 64/13,
e-mail: aap2011@ukr.net
e-mail: oteslenkoua@gmail.com,

У статті розглянуто підходи до розв'язання задачі визначення активності водія з камер встановлених у автомобілі. В статті представлений огляд основних проблем, які виникають у водія при водінні, досліджуються методи вирішення виявлених проблем. Основна частина статті зосереджена на вирішенні проблеми визначення стану водія в інтелектуальних систем допомоги водіям. Досліджуються основні парадигми вирішення цієї задачі - класифікація зображень, класифікація відеопотоку, визначення основних точок тіла водія.

Ключові слова: стан водія, система допомоги водію, камера, нейронна мережа.

The article discusses approaches to solving the problem of determining the activity of the driver from the cameras installed in the car, given the active development of intelligent driver assistance systems in recent years. The article provides an overview of the main problems that arise for the driver while driving. Main advances in autonomous driving are presented and the classification of types of autonomous vehicles is provided. Next, the methods of solving the identified problems are described. The main part of the article focuses on solving the problem of determining the state of the driver during driving. Reasons for usage of computer vision and machine learning approaches for solving this task are described. The basic paradigms of the solution of this problem - classification of images, classification of a video stream, detection of the basic points of a body of the driver on the image from the camera installed inside a car are investigated. Main ideas of every method are described. The approaches are evaluated with identification of main advantages and drawbacks of the presented methods.

Key words: driver activity, driver assistance system, camera, neural network.

Статтю представила канд.ф.-м.н., доцент Розора І.В.

Вступ

В даний час автовиробники виділяють п'ять рівнів автоматизації автономних автомобілів[1]. На нульовому рівні водій особисто виконує усі операційні завдання, такі як управління кермом, гальмування, прискорення

тощо. На п'ятому рівні автоматизації не має потреби в гальмах, кермі, так як автономна система особисто контролює всі найважливіші завдання, веде моніторинг навколишнього середовища тощо. Наразі, майже всі автовиробники використовують в своїх автомобілях перший-третій рівні

автоматомності, тобто водій відповідає за більшість важливих для безпеки функцій. Тож однією з головних задач забезпечення безпечного середовища для автономних автомобілів, та водіїв, які їх використовують є визначення стану водія та його зосередженості. Для визначення стану водія можуть використовуватися камери, встановленні в автомобілі самого водія, автомобілях інших водіїв, які рухаються назустріч даному авто, або ж камери зовнішнього спостереження за дорожнім трафіком.

Комп'ютерний зір є однією з ключових областей сучасної науки, яка бурхливо розвивається, особливо за кордоном. Як наукова дисципліна, комп'ютерний зір належить до теорії та технології створення штучних систем, які отримують інформацію у вигляді зображень. Розвиток машинного навчання та комп'ютерного зору призвів до їх використання в такій галузі як автомобілебудування. І зараз світ вже бачить перші автомобілі які можуть працювати частково автономно, з мінімальним залученням водія.

Однією з основних задач для сучасних автономних автомобілів рівня 1-3 під час руху є безпека водія, тому автоматична система управління автомобіля (інтелектуальна система допомоги водіям) повинна "бути впевнена", що водій даного автомобіля, та водії які проїжджають повз автомобіль, зосереджені на дорозі. Також за водіями слідкує зовнішня система спостереження, яка може давати сигнали водіям, які заснули, розмовляють по мобільному тощо.

Розробка інтелектуальних систем допомоги водіям (IDAS), які постійно контролюють не лише навколишнє середовище та стан транспортного засобу, а й поведінку водіїв, привертає все більше уваги у всьому світі. Вважається, що IDAS є особливо актуальним для водіїв, котрі подорожують на великі відстані, оскільки вони часто їздять в самоті. Використання інтелектуальних систем допомоги водіям, які надають важливу інформацію з поза транспортним засобом, наприклад, індикатори смуги руху та дорожні знаки, підвищують готовність водія до небезпечних ситуацій. Однак автоматичне виявлення та попередження рівня втоми та відволікання водія вважається однаково

важливим щодо запобігання ДТП. Окрім причин підвищення безпеки дорожнього руху, існують також комерційні причини встановлення систем моніторингу стану водіїв. [3]

Для визначення стану водія та ступеню його зосередженості використовується декілька підходів на основі нейронних мереж. Докладніше дані підходи будуть представлені далі.

Метод класифікації зображень

Сьогодні більшість найсучасніших рішень для розпізнавання активності водіїв використовують підхід глибокого навчання нейронних мереж[3]. Основним вдосконаленням порівняно з класичними способами є використання автоматичних характеристичних ознак отриманих з нейронних мереж, а не ручне їх створення, як в традиційних практиках комп'ютерного зору.

Один з широко популярних методів - це використання конволюційних нейронних мереж для класифікації активності водіїв[2]. Конволюційні нейронні мережі (CNN або ConvNet) - це особливий вид багатосарових нейронних мереж, розроблений для розпізнавання візуальних шаблонів безпосередньо з піксельних зображень з мінімальною попередньою обробкою.

Для цього підходу модель очікує зображення з водіями як вхідні дані, а як вихідний, буде один із заздалегідь визначених класів активності водія (Нормальне водіння, текстові повідомлення, розмова тощо). Загальна архітектура CNN для цього методу зображена нижче на рисунку 1.

Рис.1 Архітектура нейронної мережі для класифікації стану водія

Основна перевага цього підходу полягає в тому, що вже є розроблені архітектури нейронних мереж для завдань класифікації зображень, які можуть бути використані для цього підходу і дають високу точність. Найбільш підходящими архітектурами є VGG-16, Xception, ResNet[5].

У навчанні CNN часто використовується відомий алгоритм зворотного поширення похибки в архітектурі мережі. Тренінг складається з двох етапів: (i) передача навчальних даних через мережу до остаточного вихідного рівня та, нарешті, обчислення помилки чи значення функції втрати (ii) назад розповсюджується шари помилок та оновлюються ваги у відповідних шарах на основі зворотних поширюваних помилок.

Відповідна оцінка втрат для виходу з кінцевого шару - це крос-ентропійна втрата. У вихідному шарі функція крос-ентропійної втрати задається:

$$L = - \sum_{j=1}^K [t^j \log(p^j) + (1-t^j) \log(1-p^j)] \quad (1)$$

де t^j і p^j - j-й вхід і вихід відповідно, K - кількість вихідних нейронів (клас) і t^j - цільова мітка, кодована для одного класу j-го класу. Виконуючи зворотне поширення помилки, першим кроком є обчислення градієнта втрат за частинною похідною наступним чином:

$$\begin{aligned} \delta^j &= \frac{\partial L^j}{\partial p^j} \cdot \frac{\partial p^j}{\partial x^j} = \left(-\frac{t^j}{p^j} + \frac{1-t^j}{1-p^j}\right) \cdot p^j (1-p^j) \\ &= \frac{p^j - t^j}{p^j(1-p^j)} \cdot p^j(1-p^j) \\ &= p^j - t^j \end{aligned} \quad (2)$$

де δ^j - градієнт втрат у вихідному шарі, який буде поширюватися на верхній повний шар з'єднання як помилка. У l -му шарі повного з'єднання помилка δ^j $l+1$ повертається назад з верхнього $(l+1)$ -го шару.

Метод класифікації відео

Більш просунутий підхід - класифікація активності водія на основі відео[4]. Спочатку з кожного відео вибирається N кадрів. Далі до кожного обраного кадру застосовується CNN, яка заздалегідь натренована на великому датасеті

зображень. Потім витягнуті характеристичні ознаки об'єднуються і застосовуються як вхід для класифікації, що досягається додатковим рівнем нейронної мережі для прогнозування ймовірностей класу стану водія. Весь алгоритм проілюстрований на рисунку 2[5].

Рис.2 Архітектура нейронної мережі для класифікації стану водія по відео

Крім того, замість кількох звичайних CNN можна використовувати одну 3D-конволюційну нейронну мережу. 3D-конволюційна нейронна мережа очікує, що N кадрів є вхідними даними в нейронну мережу[5]. У приведеній роботі вирішується проблема маневрового прогнозування транспортних засобів за допомогою відеоданих, яка має на меті виявити активність водія на майбутнє на основі даного зразка відео. З цією метою використовується 101-шарова глибока нейронна архітектура, яка використовує 3D-згортки в поєднанні з залишковими зв'язками. Концептуально модель прогнозування активності водія складається з трьох компонентів: 1) оптична мережа вилучення оптичного потоку, 2) нейронна конволюційна мережа на основі тривимірних згортків із залишковими зв'язками для класифікації активності водія та 3) Довга короткострокова пам'ять(LSTM) для обробки відеоданих різної довжини

З часової точки зору, кожен рівень 3D нейронної мережі витягує темпоральну інформацію з вхідних кадрів або характеристичних ознак. Приклад тривимірної нейронної мережі показаний на рисунку 3[5].

Рис.3 Архітектура 3D нейронної мережі

LSTM можуть запам'ятовувати довгі часові інтервали, не втрачаючи здатності до вивчення коротких часових залежностей. Математична формулювання мережі LSTM виглядає наступним чином:

$$\begin{aligned} i_t &= \sigma(W_{ii} \cdot x_t + b_{ii} + W_{hi} \cdot h_{t-1} + b_{hi}) \\ f_t &= \sigma(W_{if} \cdot x_t + b_{if} + W_{hf} \cdot h_{t-1} + b_{hf}) \\ g_t &= \tanh(W_{ig} \cdot x_t + b_{ig} + W_{hg} \cdot h_{t-1} + b_{hg}) \\ o_t &= \sigma(W_{io} \cdot x_t + b_{io} + W_{ho} \cdot h_{t-1} + b_{ho}) \\ c_t &= f_t c_{t-1} + i_t g_t \\ h_t &= o_t \tanh(c_t) \end{aligned}$$

де i_t , f_t і o_t - ворота вводу, забуття і виведення для часу кроку t відповідно. Прихований стан h_{t-1} використовується разом із цим вхідним зразком x_t для отримання етапу g_t . Ворота вводу регулюють додавання інформації про вхідного кандидата до попереднього стану c_{t-1} , тоді як ворота забуття f_t контролюють втрату інформації про стан нейрону. h_t є виходом мережі на етапі t . W , σ відносяться до вагових матриць, зміщення та сигмоїдної функції відповідно.

Метод оцінки положення тіла та голови

Повний алгоритм для моніторингу стану водія на основі лише RGB-зображень може бути розроблений на основі оцінки положення тіла та голови. Рішення складається з середньо зваженого ансамблю згорткових нейронних мереж. Початкова CNN відповідає за виявлення ключових точок тіла водія. Для визначення верхнього скелета тіла з 13 основних точок використовуються нейронні мережі для детектування таких ключових точок. Одна із архітектура, показана на рис. 4[7], ітеративно

передбачає поля спорідненості, що кодують асоціацію "частина-частина", та достовірні карти виявлення. Наступна ітеративна архітектура прогнозування уточнює прогнози для послідовних етапів, $t \in \{1, \dots, T\}$, з проміжним наглядом на кожному етапі[7].

Рис.4 Архітектура нейронної мережі для визначення верхнього скелета тіла з 13 основних точок

У представленій архітектурі моделі рецептивне поле під час обчислення було зменшено порівняно з [7], замінивши кожне згорткове ядро 7×7 на 3 послідовні 3×3 ядра. Тоді як кількість операцій за першому слої - $2 \times 7 \times 2 - 1 = 97$, для останнього - лише 51. Додатково, вихід кожного з 3 згорткових ядер з'єднуються за принципом, подібним до підходу DenseNet. Кількість нелінійних шарів збільшується втричі, і мережа може зберігати як нижчі(прямі лінії, кути, кольори), так і вищі(положення рук, ніг, голови відносно один одного) характеристичні ознаки.

Зображення аналізується CNN (ініціалізовані перші 10 шарів VGG-19 [10] та предтренеровані), створюючи набір характеристичних карт F , які вводяться на першому етапі. На цій стадії, мережа виробляє набір полів спорідненості для кожної частини тіла.

Клас активності водія в основному розпізнається на основі оцінки напрямку голови, тому наступним кроком у цьому методі є оцінка положення та орієнтації голови. Оцінка позиції голови формулюється як багаторівнева регресійна задача до трьох кутів обертання Ейлера[9]. Для цього завдання також може бути використана CNN. Виходи нейромережі - це три кути: yaw, pitch, roll, які визначають орієнтацію голови. Архітектура представлена авторами моделі складається з двох основних частин:

магістральної мережі та трьох повністю з'єднаних шарів (по одному для кожного кута Ейлера) з використанням активації softmax. Три повністю з'єднані шари поділяють ту саму магістральну мережу. Важливо зазначити, що вихід кожного кута Ейлера з шару, що повністю з'єднується, не є одиничним значенням. Натомість це вектор, який є гіперпараметром мережі, який повинний бути підібраний. У роботі [9] автори демонструють ефективність двох різних магістральних мереж: ResNet50 та AlexNet. Вони показали, що використання ResNet50 як магістральної мережі дає кращий результат порівняно з AlexNet. Далі автори пропонують використовувати три окремих функцій втрат, по одній на кожен кут. Кожна втрата - це комбінація двох компонентів: класифікації пози та регресії позиції. Будь-яка магістральна мережа може використовуватися та доповнюватися трьома повністю сполученими шарами, які передбачають кути. Ці три повністю пов'язані шари поділяють попередні згорткові шари мережі. Ідея цього підходу полягає в тому, що, виконуючи класифікацію, використовується дуже стійкий шар softmax і крос-ентропія, таким чином мережа вчиться чітко прогнозувати позу голови. Маючи три крос-ентропійні втрати, по одній на кожен кут Ейлера, отримують три сигнали, які передаються в мережу, що покращує навчання. Потім в мережу додаються втрати

регресії, а саме середньоквадратичну втрату помилок, щоб покращити кінцеві прогнози. В кінці отримують три кінцеві витрати, по одній на кожен кут, і кожна є лінійною комбінацією як відповідної класифікації, так і регресійних втрат.

Висновки

Було представлено підходи до відтворення характеристик стану водія в автомобілі з рівнями автоматизації 1-3 за допомогою нейронних мереж. На даний момент є декілька підходів до вирішення цієї задачі. Один з підходів це використання відеопотоку з камер встановлених в автомобілі, або зовні для тренування нейронної мережі для покадрової класифікації зображень. Також крім окремих кадрів можуть бути класифіковані короткі відеофрагменти. Для цього підходу використовуються нейронні мережі на основі 3D згорток. Останній метод полягає в визначенні ключових точок на тілі водія, визначення ключових точок голови водія, та кути повороту голови. Наступним етапом ці ключові точки використовуються для класифікації стану водія за допомогою нейронної мережі.

Список використаних джерел

1. Yan C. Vision based driver behaviour analysis [Електронний ресурс] / Chao Yan // University of Liverpool. – 2016. – Режим доступу до ресурсу: https://cgi.csc.liv.ac.uk/~frans/CurrentResearch/Thesis/yan_March2016.pdf.
2. Tamas V. Real-Time Distracted Drivers Detection Using Deep Learning [Електронний ресурс] / V. Tamas, V. Maties // IET Intelligent Transport Systems (IET INTELL TRANSP SY). – 2019. – Режим доступу до ресурсу: https://www.researchgate.net/publication/326740203_Real-time_Detection_of_Distracted_Driving_based_on_Deep_Learning.

References

1. Yan C. Vision based driver behaviour analysis [Electronic resource] / Chao Yan // University of Liverpool. – 2016. – Available from https://cgi.csc.liv.ac.uk/~frans/CurrentResearch/Thesis/yan_March2016.pdf.
2. Tamas V. Real-Time Distracted Drivers Detection Using Deep Learning [Electronic resource] / V. Tamas, V. Maties // IET Intelligent Transport Systems (IET INTELL TRANSP SY). – 2019. – Available from: https://www.researchgate.net/publication/326740203_Real-time_Detection_of_Distracted_Driving_based_on_Deep_Learning.

3. Gebert P. End-to-end Prediction of Driver Intention using 3D Convolutional Neural Networks [Електронний ресурс] / P. Gebert, A. Roitberg, M. Haurilet // Conference: 2019 IEEE Intelligent Vehicles Symposium (IV). – 2019. – Режим доступу до ресурсу: https://www.researchgate.net/publication/335497206_End-to-end_Prediction_of_Driver_Intention_using_3D_Convolutional_Neural_Networks.

4. Eraqi H. Driver Distraction Identification with an Ensemble of Convolutional Neural Networks [Електронний ресурс] / Hesham Eraqi // Journal of Advanced Transportation. – 2019. – Режим доступу до ресурсу: <https://arxiv.org/abs/1901.09097>.

5. ation using Part Affinity Fields [Електронний ресурс] / Z. Cao, T. Simon // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – Режим доступу до ресурсу: <https://arxiv.org/abs/1611.08050>.

6. Alioua N. Driver head pose estimation using efficient descriptor fusion [Електронний ресурс] / N. Alioua, A. Amine, A. Rogozan // EURASIP Journal on Image and Video Processing. – 2016. – Режим доступу до ресурсу: <https://jivp-urasipjournals.springeropen.com/articles/10.1186/s13640-016-0103-z>.

7. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields [Електронний ресурс] / Z.Cao, G. Hidalgo Martinez, T. Simon, S. Wei // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2019. – Режим доступу до ресурсу: <https://arxiv.org/abs/1812.08008>.

8. Simonyan K. Very Deep Convolutional Networks for Large-Scale Image Recognition [Електронний ресурс] / K. Simonyan, A. Zisserman // ICLR 2015. – 2014. – Режим доступу до ресурсу: <https://arxiv.org/abs/1409.1556>.

9. Ruiz N. Fine-Grained Head Pose Estimation Without Keypoints [Електронний ресурс] / N. Ruiz, E. Chong, J. M. Rehg // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – Режим доступу до ресурсу: <https://arxiv.org/abs/1710.00925>.

3. Eraqi H. Driver Distraction Identification with an Ensemble of Convolutional Neural Networks [Electronic resource] / Hesham Eraqi // Journal of Advanced Transportation. – 2019. – Available from: <https://arxiv.org/abs/1901.09097>.

4. Eraqi H. Driver Distraction Identification with an Ensemble of Convolutional Neural Networks [Electronic resource] / Hesham Eraqi // Journal of Advanced Transportation. – 2019. – Available from: <https://arxiv.org/abs/1901.09097>.

5. Cao Z. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields [Electronic resource] / Z. Cao, T. Simon // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – Available from: <https://arxiv.org/abs/1611.08050>.

6. Alioua N. Driver head pose estimation using efficient descriptor fusion [Electronic resource] / N. Alioua, A. Amine, A. Rogozan // EURASIP Journal on Image and Video Processing. – 2016. – Available from: <https://jivp-urasipjournals.springeropen.com/articles/10.1186/s13640-016-0103-z>.

7. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields [Electronic resource] / Z.Cao, G. Hidalgo Martinez, T. Simon, S. Wei // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2019. – Available from: <https://arxiv.org/abs/1812.08008>.

8. Simonyan K. Very Deep Convolutional Networks for Large-Scale Image Recognition [Electronic resource] / K. Simonyan, A. Zisserman // ICLR 2015. – 2014. – Available from: <https://arxiv.org/abs/1409.1556>.

9. Ruiz N. Fine-Grained Head Pose Estimation Without Keypoints [Electronic resource] / N. Ruiz, E. Chong, J. M. Rehg // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – Available from: <https://arxiv.org/abs/1710.00925>.

Надійшла до редколегії: 20.01.2020