

**Боділовський О.К.**

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

## ВИБІР ОБЛАСТІ ІНТЕРЕСУ В РАЗІ БЕЗКОНТАКТНОГО МОНІТОРИНГУ ПАРАМЕТРІВ ДИХАННЯ З ВИКОРИСТАННЯМ СЕМАНТИЧНОЇ СЕГМЕНТАЦІЇ

У статті запропоновано метод вибору області інтересу при безконтактному моніторингу дихання за допомогою камери. Запропонований метод базується на використанні семантичної сегментації тіла людини з метою визначення ділянки, яка опосередковано бере участь у диханні. Метод полягає в автоматичному визначенні ділянки торсу людини в кадрі та отриманні кривої дихання за допомогою методу оптичного потоку. Отриманий сигнал не міститиме сторонніх рухів об'єктів фону, які могли спотворювати якість моніторингу дихання. Для визначення області інтересу використовувався модель, що дає змогу локалізувати характерні ділянки тіла людини (ділянка голови, торсу, плечей, передпліч тощо) для кадрів відеопослідовності. У подальшому сегментований кадр використовувався під час визначення області інтересу, для якої проводилася оцінювання оптичного потоку. Отримана в результаті крива оптичного потоку відображає характеристики дихання людини в часовій області. Запропонований алгоритм дає змогу, зокрема, оцінити частоту дихання, а також складніші параметри дихання, що мають діагностичну цінність під час моніторингу стану пацієнта. Продемонстровано отримання кривої дихання з відеозапису з вибором області інтересу за допомогою семантичної сегментації. Розрахована за допомогою алгоритму крива дихання не має в складі складників, пов'язаних зі стороннім рухом, наявним у кадрах на відеозаписі, а містить складники, пов'язані з дихальними рухами об'єкта. Перевагою цього методу є те, що можна оцінити розташування об'єкта дослідження в кадрі та використовувати інформацію про розташування інших частин тіла з метою покращення якості аналізу. Недоліком використаної моделі є неможливість її використання в мобільних пристроях і в пристроях, що не мають у складі потужних графічних процесорів. Тому актуальним завданням для цього алгоритму є використання моделей, що можуть бути запущені на мобільному пристрої та дозволятимуть моніторинг дихання пацієнта в режимі реального часу. Цей метод може застосовуватися в безконтактних системах моніторингу дихання в медичних закладах, домашньому моніторингу, на транспорті тощо.

**Ключові слова:** безконтактний моніторинг дихання, семантична сегментація, параметри дихання, вибір області інтересу, оптичний потік.

**Постановка проблеми.** Моніторинг дихання є важливим аспектом у задачах усунення артефактів руху при комп'ютерній томографії [1], магнітно-резонансній томографії [2], променевої терапії [3], у неонатальних застосуваннях [4], у приладах моніторингу стану людей похилого віку [5], а також у приладах моніторингу для автомобільної, залізничної та авіаційної галузі, контролю на транспорті. Останнім часом до безконтактних методів прикута все більша увага у зв'язку з посиленням заходів безпеки при пандеміях, зокрема, для виявлення хворих в аеропортах, вокзалах тощо [6].

**Аналіз останніх досліджень і публікацій.** Значні зусилля спрямовані на проектування безконтактних систем моніторингу дихання в реальному часі. Відсутність прямого контакту між пацієнтом і пристроєм контролю не завдає дис-

комфорту й розширює сфери можливого застосування. Є багато підходів до безконтактного моніторингу дихання, таких як ультразвуковий метод, радарний метод, метод на основі емісійної ЕКГ тощо. Але найбільш перспективними вважаються методи, засновані на аналізі відеозображення пацієнта, яке може бути отримане за допомогою камер, що пов'язано з широкою доступністю камер, їх низькою вартістю й задовільною якістю отриманого відеосигналу. Для підвищення точності моніторингу параметрів дихання важливим завданням є визначення області інтересу на відео, у якій будуть наявні лише ті ділянки тіла, що рухаються в процесі дихання, а також ця ділянка кадру має бути вільна від артефактів. У роботі запропоновано метод визначення області інтересу з використанням семантичної сегментації, яка знайшла широке застосування в області

комп'ютерного зору, починаючи від сегментації в мікроскопії й закінчуючи автономними автомобілями, та обґрунтовано її застосування.

**Постановка завдання.** Метою статті є розгляд методу вибору області інтересу при безконтактному моніторингу дихання за допомогою камери.

**Виклад основного матеріалу дослідження.**

**1. Метод вибору області інтересу при безконтактному моніторингу дихання**

При безконтактному моніторингу дихання за допомогою камери та на основі оптичного потоку визначним фактором точності методу є вибір ділянки кадру для моніторингу та відсутність різноманітних артефактів і завад, адже необхідно відслідковувати майже непомітні неозброєним оком коливання грудини. Сьогодні більшість підходів на безконтактного моніторингу дихання за допомогою однієї відеокамери, які описані в публікаціях [7–10], мають такий алгоритм: визначення області інтересу, що може бути виконана за рахунок установлення камери, та об'єктива на відстані, при якій у кадр потраплятиме лише область інтересу (грудина), ручне виділення області інтересу з подальшим її аналізом, а можливо, й автоматичним відстеженням, також застосовують різноманітні підходи для вибору характерних точок як підсилення відео на основі методу Ейлера (Eulerian Video Magnification) [11]. Після вибору області інтересу застосовують метод оптичного потоку [12] для визначення швидкості руху пікселів у часі між кадрами відеопослідовності. Так отримується крива, яка відповідає швидкості руху характерних точок в області інтересу, що корелює з дихальними рухами грудини людини. У подальшому отриманий сигнал обробляють з метою знаходження частоти дихання, застосовуючи або підходи в частотній області з метою визначення основної частоти, або аналіз у часовій області з метою визначення піків дихальної кривої й розрахунку частоти коливань. Зазвичай аналіз проводять із застосуванням фільтрації або методу головних компонент PCA [13].

Розглянемо переваги та недоліки описаних вище підходів до вибору області інтересу. Мето-

дика налаштування камери так, щоб у кадр потрапила лише область інтересу, має незаперечну перевагу через відсутність можливих артефактів, спричинених потраплянням у кадр сторонніх об'єктів на фоні, що може спотворити результати аналізу. Тому такий метод може мати обмежену функціональність і застосований лише для пацієнтів, що перебувають нерухомо на лікарняному ліжку, або ж до немовлят, які перебувають в інкубаційному боксі. На відміну від попереднього підходу, ручний вибір області інтересу позбавлений недоліку встановлення камери лише певним чином, а може бути використаний під час моніторингу за допомогою багатьох відеокамер, включаючи веб-камери, а також камери, що встановлені в мобільні телефони чи ноутбуки. Ручний метод можна зробити напівавтоматичним, додавши до алгоритму трекінг, коли оператор вказує вручну область моніторингу, а надалі алгоритм буде «супроводжувати» цю область від кадру до кадру на відео за певними ознаками (колір, структура, характерні точки тощо). Недоліком напівавтоматичного методу можна вважати необхідність оператору вказати область інтересу на початку моніторингу, а також ненадійність методів трекінгу, які можуть «губити» область інтересу, що знову ж таки потребуватиме залучення оператора.

Основними вимогами до вибору області інтересу є точне виділення ділянки тіла людини, яка виконує дихальні рухи. Область торсу має найбільший інтерес під час моніторингу параметрів дихання, адже дихальні найбільш помітні саме в цій області. Залежно від типу дихання, черевного (коли задіяні більшою мірою м'язи діафрагми та живота) або грудного (задіяні м'язи грудини та діафрагми), найбільш амплітудні рухи зосереджені в ділянці торсу. Також важливою є розроблення автоматичного методу вибору області інтересу, щоб виключити необхідність участі оператора в процесі моніторингу.

У роботі пропонується використання семантичної сегментації для визначення області інтересу. Це дасть змогу уникнути недоліків, пов'язаних із

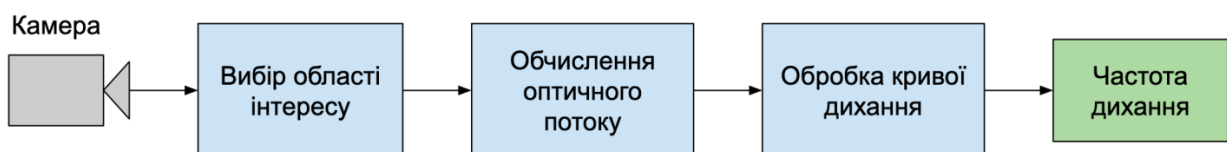


Рис. 1. Схема алгоритму безконтактного моніторингу дихання

наявними методами вибору області інтересу при безконтактному моніторингу параметрів дихання.

## 2. Семантична сегментація

Сегментація зображень – це сукупність методів комп’ютерного зору, що мають на меті розбиття зображення на сукупності пікселів, які відповідають одному об’єкту чи його частині. У свою чергу, семантична сегментація присвоює кожному пікселю зображення мітку, яка вказує на належність до певного класу. На рис. 2 зображено приклад семантичної сегментації, де групи пікселів, що відповідають за силуети автомобілів, людей, дороги, тротуару тощо, присвоєно відповідному класу та зафарбовано відповідним для кожного класу кольором.

Методи сегментації мають тривалу історію, їх умовно можна розділити на класичні або традиційні, методи сегментації за допомогою глибоких нейронних мереж. До традиційних можна зарахувати сегментацію за кольором пікселів зображення, а також сегментацію за допомогою підходів, що засновані на виділенні різноманітних властивостей об’єктів на зображенні за допомогою harris corners [15], histogram of oriented gradients (HOG) [16], дескрипторів scale-invariant feature transform (SIFT) [17], Speeded-Up Robust Features (SURF) [18] тощо, а також підходів, Bag-of-visual-words (BOV) [19], Textons [20] тощо. На цьому етапі розвитку алгоритмів сегментації традиційні методи вже давно програють у якості алгоритмам на основі нейронних мереж.

Напрямок методів семантичної сегментації на основі глибоких нейронних мереж також має кілька підходів, які можна розділити так [21]:

1. Повнозв’язкові згорткові нейронні мережі (Fully convolutional networks).

2. Згорткові нейронні мережі з графічними моделями (Convolutional models with graphical models).

3. Мережі на основі енкодер-декодер архітектури (Encoder-decoder based models).

4. Багатомасштабні та пірамідальні мережі (Multi-scale and pyramid network based models).

5. Мережі, засновані на R-CNN архітектурі (для інстанс сегментації).

6. Моделі, засновані на розріджених згортках (dilated convolution), а також сімейство моделей DeepLab.

7. Моделі, засновані на рекурентних нейронних мережах.

8. Моделі з використанням уваги (Attention-based).

9. Генеративні змагальні мережі.

10. Згорткові моделі з використанням активного контуру (Convolutional models with active contour models).

11. Інші моделі.

Для тренування нейронних мереж і порівняння різноманітних алгоритмів і підходів використовують стандартизовані набори даних: PASCAL VOC [22], MS COCO [23], ADE20k [24; 25], Cityscapes [14] тощо, а також стандартизовані метрики якості роботи алгоритмів, такі як піксельна точ-



Рис. 2. Приклад результату семантичної сегментації зображення з набору даних Cityscapes [14]

ність (Pixel accuracy) або середня піксельна точність (mPA), коефіцієнт Жаккара (Intersection over Union(IoU)) або його середнє значення для кожного класу (mean-IoU), F1-коефіцієнт, DICE коефіцієнт тощо.

Як приклад наведено таблицю порівняння методів сегментації за метрикою mean-IoU для набору даних PASCAL VOC 2012 test [22].

### 3. Вибір області інтересу за допомогою семантичної сегментації

Пропонується використовувати вже наявний набір даних Pascal-Person-Part [27]. Цей набір даних побудований на основі вже наявного набору PASCAL VOC, до якого додані нові файли розмітки, що дають можливість натренувати модель на сегментацію семи класів: голова, плече, передпліччя, торс, верхня частина ніг, нижня частина ніг, а також фон. Цей набір класів дає змогу з високою точністю й автоматично визначати область інтересу для моніторингу параметрів дихання залежно від обраної моделі та наявних обчис-

лювальних ресурсів, уточнювати область інтересу в кожному кадрі або ж раз на декілька кадрів, а також цей підхід дає можливість усунути артефакти, пов'язані з перекриттям області інтересу кінцівками пацієнта.

Для проведення експерименту використану модель, представлену в роботі [28]. Модель має архітектуру під назвою Augmented Context Embedding with Edge Perceiving (A-CE2P). На рисунку 3б наведено результат вибору області інтересу порівняно з ручним методом вибору (рисунок 3а). Як видно з рисунка, в область інтересу, обрану вручну, потрапляє рука об'єкта дослідження, що може вплинути на якість аналізу параметрів дихання через артефакти руху, тоді як область інтересу, обрана за допомогою сегментації, включає лише торс людини (зелений колір), що дає змогу більш точно отримати криву дихання.

У роботі використана модель [28] з метою визначення області інтересу з подальшим

Таблиця 1

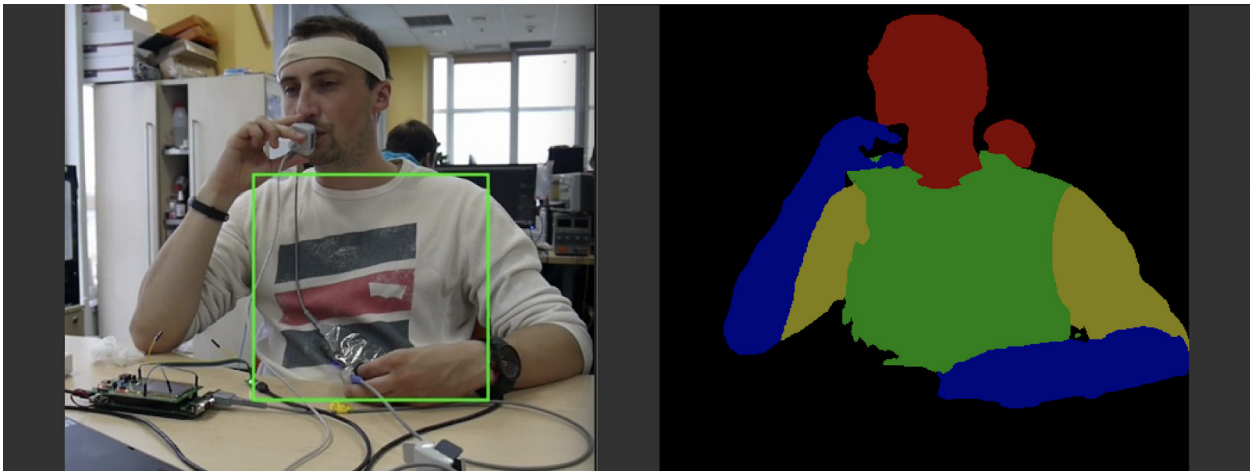
Методи семантичної сегментації та їх ефективність [26]

| RANK | METHOD                                      | MEAN IOU | EXTRA TRAINING DATA | PAPER TITLE   | YEAR |
|------|---|----------|---------------------|---|------|
| 1    | DeepLabv3+ (Xception-65-JFT)                | 89.0%    | ✓                   | Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation   | 2018 |
| 2    | ExFuse (ResNeXt-131)                        | 87.9%    | ✓                   | ExFuse: Enhancing Feature Fusion for Semantic Segmentation                          | 2018 |
| 3    | DPC   | 87.9%    | ✓                   | Searching for Efficient Multi-Scale Architectures for Dense Image Prediction        | 2018 |
| 4    | CFNet (ResNet-101)                          | 87.2%    | ✓                   | Co-Occurrent Features in Semantic Segmentation                                      | 2019 |
| 5    | DeepLabv3-JFT                               | 86.9%    | ✓                   | Rethinking Atrous Convolution for Semantic Image Segmentation                       | 2017 |
| 6    | DCNAS                                       | 86.9%    | ×                   | DCNAS: Densely Connected Neural Architecture Search for Semantic Image Segmentation | 2020 |
| 7    | Smooth Network with Channel Attention Block | 86.2%    | ×                   | Learning a Discriminative Feature Network for Semantic Segmentation                 | 2018 |
| 8    | SANet (pretraining on COCO dataset)         | 86.1%    | ✓                   | Squeeze-and-Attention Networks for Semantic Segmentation                            | 2019 |
| 9    | EncNet                                      | 85.9%    | ✓                   | Context Encoding for Semantic Segmentation  | 2018 |
| 10   | PSPNet                                      | 85.4%    | ✓                   | Pyramid Scene Parsing Network   | 2016 |

аналізом параметрів дихання, що оцінені на основі запропонованої моделлю області. На рисунку 4 наведено приклад кривої дихання, що оцінена з області інтересу.

Використана для вибору області інтересу модель має недолік, через те що на визначення області інтересу потрібні значні обчислювальні ресурси й у реальному часі робота цієї моделі можлива лише за наявності потужного графіч-

ного прискорювача. Щоб уможливити аналіз з автоматичним вибором області інтересу на основі семантичної сегментації в реальному часі, необхідно використати інші архітектури нейронних мереж, а також адаптувати їх до роботи на мобільних пристроях. Це дасть змогу виконувати точніший моніторинг параметрів дихання на більшій кількості розповсюджених пристроїв.



а) б)  
Рис. 3. Порівняння вручну обраної області інтересу та області, отриманої за допомогою семантичної сегментації

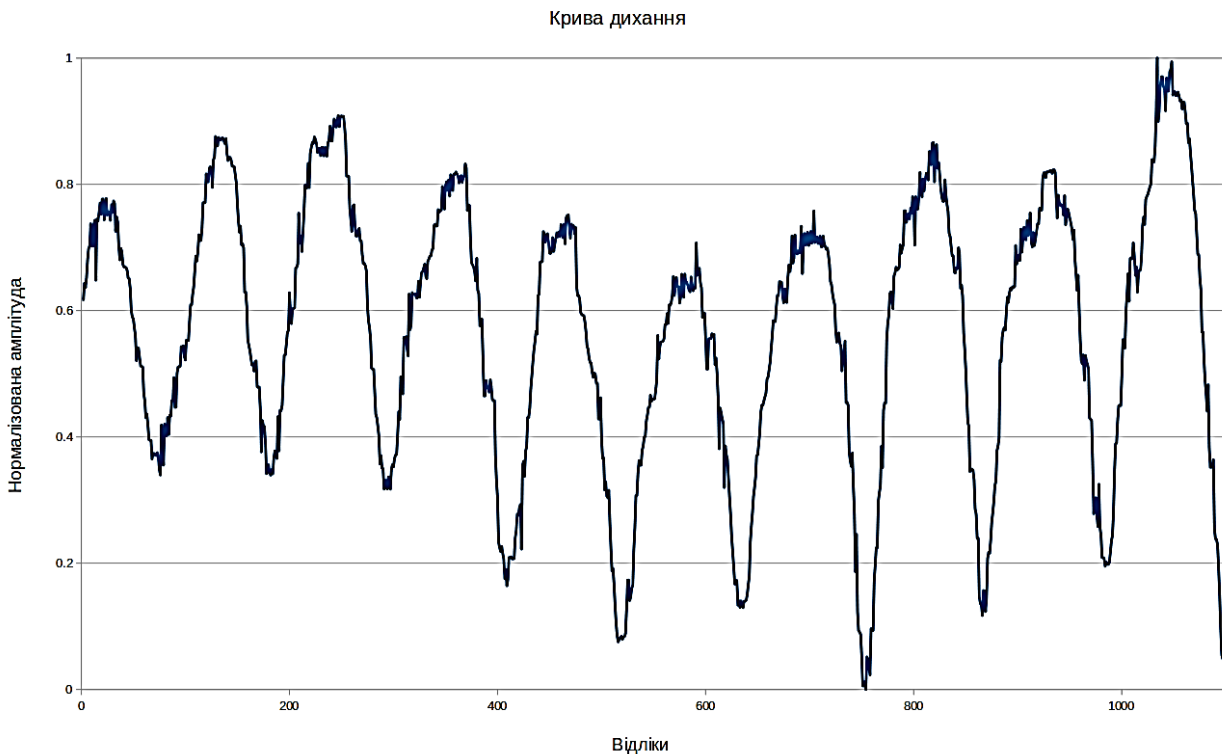


Рис. 4. Приклад кривої дихання з області інтересу, визначеної за допомогою семантичної сегментації

**Висновки.** У роботі запропоновано підхід до автоматичного вибору області інтересу при безконтактному моніторингу параметрів дихання за допомогою камери на основі семантичної сегментації тіла людини. Продемонстровано використання нейронної мережі для визначення області інтересу, а також отримано криву дихання, що обчислена на основі аналізу відео із цієї області.

Визначено основний недолік нейронної мережі, яка використовувалася для сегментації, а саме обмеження використання її в реальному часі на пристроях без потужного графічного прискорювача. Для цього вже розпочато роботу над створенням моделі, що не матиме цього недоліку й працюватиме на мобільному пристрої або ноутбуку з прийнятною точністю.

#### Список літератури:

1. Jones K.R. A respiration monitor for use with CT body scanning and other imaging techniques. *Br. J. Radiol.* 1982. Vol. 55. № 655. P. 530–533.
2. Magnetic resonance imaging with respiratory gating: techniques and advantages / R. Ehman, M. McNamara, M. Pallack, H. Hricak, and C. Higgins. *Am. J. Roentgenol.* 1984. Vol. 143. № 6. P. 1175–1182.
3. Li T., Geng J., Li S. Automatic respiration tracking for radiotherapy using optical 3D camera. 2013. P. 861804–861804.
4. Neonatal Non-contact Respiratory Monitoring based on Real-time Infrared Thermography / A.K. Abbas, K. Heimann, K. Jergus, T. Orlikowsky, S. Leonhardt. *Biomed. Eng. OnLin.* 2011. Vol. 10. № 1. P. 93.
5. Aoki H., Koshiji K. Non-contact Respiration Monitoring Method for Screening Sleep Respiratory Disturbance Using Slit Light Pattern Projection,” in *World Congress on Medical Physics and Biomedical Engineering 2006*. R. Magjarevic and J. H. Nagel, Eds. Springer Berlin Heidelberg, 2007. P. 680–683.
6. Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner / Y. Wang, M. Hu, Q. Li, X.-P. Zhang, G. Zhai, N. Yao. *ArXiv200205534 Cs Eess*, Feb. 2020, Accessed: Mar. 23, 2020. URL: <http://arxiv.org/abs/2002.05534>.
7. Nakajima K., Osa A., Miike H. A method for measuring respiration and physical activity in bed by optical flow analysis. *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1997*. 1997. Vol. 5. P. 2054–2057.
8. Nakajima K., Matsumoto Y., Tamura T. Development of real-time image sequence analysis for evaluating posture change and respiratory rate of a subject in bed. *Physiol. Meas.* 2001. Vol. 22. № 3. № 21.
9. Kuo Y.-M., Lee J.-S., Chung P. A Visual Context-Awareness-Based Sleeping-Respiration Measurement System. *IEEE Trans. Inf. Technol. Biomed.* 2010. Vol. 14. № 2. P. 255–265.
10. Real-time vision based respiration monitoring system / K.S. Tan, R. Saatchi, H. Elphick, D. Burke. *2010 7th International Symposium on Communication Systems Networks and Digital Signal Processing (CSNDSP)*. 2010. P. 770–774.
11. Eulerian video magnification and analysis / N. Wadhwa et al. *Commun. ACM.* 2016. Vol. 60. № 1. P. 87–95.
12. Horn B.K.P., Schunck B.G. Determining optical flow. *Artif. Intell.* 1981. Vol. 17. № 1–3. P. 185–203.
13. Wold S., Esbensen K., Geladi P. Principal component analysis. *Chemom. Intell. Lab. Syst.* 1987. Vol. 2. № 1. P. 37–52.
14. The Cityscapes Dataset for Semantic Urban Scene Understanding / M. Cordts et al. presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, P. 3213–3223. URL: [http://openaccess.thecvf.com/content\\_cvpr\\_2016/html/Cordts\\_The\\_Cityscapes\\_Dataset\\_CVPR\\_2016\\_paper.html](http://openaccess.thecvf.com/content_cvpr_2016/html/Cordts_The_Cityscapes_Dataset_CVPR_2016_paper.html) (accessed: May 15, 2020).
15. Konstantinos G. Derpanis. The harris corner detector. York University, Oct. 2004. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.482.1724&rep=rep1&type=pdf> (accessed: May 15, 2020).
16. Dalal N., Triggs B. Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Jun. 2005. Vol. 1. P. 886–893.
17. Lowe D.G. Object recognition from local scale-invariant features. *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Sep. 1999. Vol. 2. P. 1150–1157.
18. Speeded-Up Robust Features (SURF) / H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. *Comput. Vis. Image Underst.* 2008. Vol. 110. № 3. P. 346–359.
19. Visual Categorization with Bags of Keypoints / G. Csurka, C. R. Dance, L. Fan, J. Willamowski, C. Bray. P. 16.
20. What are Textons? / S.-C. Zhu, C. Guo, Y. Wang, and Z. Xu. *Int. J. Comput. Vis.* 2005. Vol. 62. № 1. P. 121–143, Apr. 2005.
21. [2001.05566] Image Segmentation Using Deep Learning: A Survey. URL: <https://arxiv.org/abs/2001.05566> (accessed: May 15, 2020).
22. The Pascal Visual Object Classes (VOC) Challenge / International Journal of Computer Vision. URL: <https://dl.acm.org/doi/10.1007/s11263-009-0275-4> (accessed: May 15, 2020).

23. Microsoft COCO: Common Objects in Context / T.-Y. Lin et al. *Computer Vision – ECCV 2014*. Cham, 2014. P. 740–755.
24. Scene Parsing through ADE20K Dataset / B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, A. Torralba. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.
25. Semantic understanding of scenes through the ade20k dataset / B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, A. Torralba. *ArXiv Prepr. ArXiv160805442*, 2016.
26. Papers with Code – PASCAL VOC 2012 test Leaderboard. URL: <https://paperswithcode.com/sota/semantic-segmentation-on-pascal-voc-2012> (accessed: May 15, 2020).
27. Detect What You Can: Detecting and Representing Objects Using Holistic Models and Body Parts / X. Chen, R. Mottaghi, X. Liu, S. Fidler, R. Urtasun, A. Yuille. *2014 IEEE Conference on Computer Vision and Pattern Recognition*. Jun. 2014. P. 1979–1986.
28. Self-Correction for Human Parsing / P. Li, Y. Xu, Y. Wei, Y. Yang. *ArXiv191009777 Cs Eess*. Oct. 2019. URL: <http://arxiv.org/abs/1910.09777> (accessed: May 15, 2020).

### **Bodilovskyi O.K. SELECTION OF REGION OF INTEREST IN NON-CONTACT MONITORING OF RESPIRATION PARAMETERS USING SEMANTIC SEGMENTATION**

*The article proposes a method for selection of the region of interest in camera-based non-contact respiration monitoring solutions. The proposed method is based on the semantic segmentation of the human body and performs a selection of the area that is corresponded to respiration movements. The method is automatically determining the area of the human torso in the frame and obtaining the respiration curve using the optical flow method. The received signal will not contain artifact movements of background objects that could decrease the quality of respiratory monitoring. A model was used for localization of the human body parts (head, torso, shoulders, forearms, etc.) in the video. After that the segmented frame was used to selection of the region of interest for the optical flow evaluation. The resulting optical flow curve reflects the characteristics of human respiration in the time domain. The proposed algorithm allows us to estimate the respiratory rate, as well as more complex respiratory parameters that have diagnostic value for the patient's monitoring. The article demonstrated an estimation of the respiration curve from a video with a region of interest selection using semantic segmentation. The curve estimated using an algorithm does not include components related to artifacts movements that were presented in the frames but contain components related to the respiratory movements of the object. The advantage of this method is the assessment of the object location in the frame as well as other parts of the body to improve the quality of the respiratory analysis. The disadvantage of this model is low performance in mobile devices and devices that do not include powerful GPUs. Therefore, an important task for this algorithm is improvements of the model performance that allow us to run the model on a mobile device and perform real-time monitoring of the patient's respiration. This method can be used in non-contact respiration monitoring systems in hospitals, for home monitoring, for respiration monitoring on public transportation, etc.*

**Key words:** non-contact respiration monitoring, semantic segmentation, respiration parameters, region of interest selection, optical flow.